





A;Reference number: JC6552; MUID:98245940; PMID:9583949  
A;Accession: JC6552  
A;Molecule type: mRNA  
A;Residues: 1-1015 <CZE>  
A;Cross-references: GB:AF023910; GB:U63217; NID:g2642492; PIDN:AAC14193.1; PID:g2642493  
C;Genetics:  
A;Gene: top1  
C;Superfamily: eukaryotic type I DNA topoisomerase  
K;Keywords: isomerase  
F;974/Active site: Tyr #status predicted

Query Match                42.2%; Score 70; DB 2; Length 1015;  
Best Local Similarity     42.5%; Pred.No. 5.9; Matches 18; Conservative 6; Mismatches 8; Indels 10; Gaps 1;

QY      2 KKVAKKEKAAYKKYE-----AKAAAEAAAEEARAY 33  
         ||| ||| ||| ||| :||| |:||||::|  
DB     293 KKEVKEDTAKOVKVEKVETPKTKPAKRKAESSESDP 334

RESULT 11

T34081  
hypothetical protein C02F12.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34081  
R;Miller, N.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid C02F12.  
A;Reference number: Z21473  
A;Status: Preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1130 <ML>  
A;Cross-references: EMBL:UA1545; PIDN:AAA83190.1; CESP:C02F12.7  
C;Genetics:  
A;Gene: CSP:C02F12.7  
A;Introns: 11/2; 118/1; 156/2; 239/2; 308/2; 398/1; 458/3; 591/1; 640/3; 756/2; 790/2; 8

Query Match                42.2%; Score 70; DB 2; Length 1130;  
Best Local Similarity     51.4%; Pred.No. 6.4; Matches 18; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY      6 KKEKAARKAYKKEAKAKA-----AEAAREEAAYE 34  
         |:||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
DB     1016 KREFAEKRTKEAKEKEKEKEKEKEAKEAKE 1050

RESULT 12

JV0057  
tolA protein - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002  
C;Accession: JV0057; B64810  
R;Levengood, S.K.; Webster, R.E.  
J. Bacteriol. 171, 5600-5609, 1989  
A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their products  
A;Reference number: JV0057; MUID:90078104; PMID:2687247  
A;Accession: JV0057  
A;Molecule type: DNA  
A;Residues: 1-421 <LEV>  
A;Cross-references: GB:M28232; NID:gi48018; PIDN:AAA24683.1; PID:gi48019  
A;Experimental source: strain JM105  
A>Note: the authors translated the initiation codon GTG for residue 1 as Val  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen,  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: B64810  
A>Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-421 <BLAT>





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 7.97468 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-1  
Perfect score: 166  
Sequence: 1 AKYAKKEKAAYKKAKEAKAEEAAKEAYEA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	84	50.6	384	1	TMPE TREPH
2	76	45.8	347	1	TOLA_PSEAE
3	74	44.6	372	1	TOLA_HAEIN
4	72	43.4	210	1	H1_LYTPI
5	71.5	43.1	208	1	DBH MYCSC
6	71.5	43.1	211	1	H1B_STRPU
7	70	42.2	1130	1	VL17_CAEEL
8	69.5	41.9	421	1	TOLA_ECOLI
9	68.5	41.3	87	1	H1C6_TRYCR
10	68.5	41.3	97	1	H1C2_TRYCR
11	68.5	41.3	101	1	ASR_ENTCL
12	66.5	40.1	217	1	H1G_STRPU
13	66.5	40.1	568	1	ZRF1_HUMAN
14	66	39.8	1391	1	MST2_DROHY
15	65.5	39.5	214	1	DBH MYCTU
16	65.5	39.5	1528	1	SPAA_STRDO
17	64.5	38.9	185	1	H1D_STRPU
18	64.5	38.9	205	1	DBH MYCBO
19	64	38.6	80	1	H161_TRYCR
20	64	38.6	90	1	H162_TRYCR
21	64	38.6	233	1	H11_GLYBA
22	64	38.6	240	1	H12_VOLCA
23	63.5	38.3	514	1	ZRF1_MOUSE
24	63.5	38.3	707	1	HS88_NEUCR
25	63	38.0	248	1	H1_PARAN
26	63	38.0	260	1	H11_VOLCA
27	62.5	37.7	232	1	H1A_CHITE
28	62.5	37.7	349	1	RS6_AEDAL
29	62	37.3	206	1	H1_ONCMY
30	62	37.3	237	1	H1E_CHITE
31	62	37.3	244	1	H1O_CHITH
32	62	37.3	898	1	IF2_VIBCH
33	61.5	37.0	233	1	H11_GLYSA

## RESULT 1

ID	TMPE TREPH	STANDARD;	PRT;	384 AA.
AC	P29720;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Treponema membrane protein B precursor (Antigen tmpB).			
GN	TMPE.			
OS	Treponema phagedenis.			
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.			
OX	NCBI_TaxID=162;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Kazan 5;			
RX	MEDLINE=91372983; PubMed=1894368;			
RA	Yelton D.B., Limberger R.J., Curci K., Malinosky-Rummell F.,			
RA	Slivienky L., Scholls L.M., van Embden J.D., Charon N.W.;			
RT	Treponema phagedenis encodes and expresses homologs of the Treponema			
RL	pallidum tmpA and tmpB proteins.";			
RL	Infect. Immun. 59:3685-3693(1991).			
CC	-!- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR			
CC	LARGE MOLECULES.			
CC	-!- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.			
CC	-!- SIMILARITY: TO TMPE OF T.PALLIDUM.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M58563; AAA27480.1; -			
DR	PIR; B43592; B43592.			
KW	Antigen; Outer membrane; Repeat; Signal.			
FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	384	TREPONEMAL MEMBRANE PROTEIN B.
FT	DOMAIN	151	235	17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR] -
FT	REPEAT	151	155	[ED].
FT	REPEAT	156	160	1-1.
FT	REPEAT	161	165	1-2.
FT	REPEAT	166	170	1-3.
FT	REPEAT	171	175	1-4.
FT	REPEAT	176	180	1-5.
FT	REPEAT	181	185	1-6.
FT	REPEAT	186	190	1-7.
FT	REPEAT	191	195	1-8.
FT	REPEAT	196	200	1-9.
FT	REPEAT	201	205	1-10.
FT	REPEAT	206	210	1-11.
FT	REPEAT	211	215	1-12.
FT	REPEAT	216	220	1-13.
FT	REPEAT	221	225	1-14.
FT	REPEAT	226	230	1-15.

## ALIGNMENTS



Query Match 43.4%; Score 72; DB 1; Length 210;  
Best Local Similarity 59.0%; Pred. No. 0.57;  
Matches 23; Conservative 3; Mismatches 9; Indels 4; Gaps 2;

QY 1 AKKYAKK--EKAACK-AVKEAKAKAAEAAAEAAEAA 35  
DB 157 AKKAACKPAKKAACKPAKKAACKPAKKAACKPAKKAACKPA 195

RESULT 5  
DBH MYCSM STANDARD; PRT; 208 AA.  
AC Q9ZHC5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).  
GN HUP OR HLP.  
OS Mycobacterium smegmatis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1772;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 700084 / mc(2)155;  
RX MEDLINE=91110209; PubMed=9894918;  
RA Lee B.H., Murugasu-Oei B., Dick T.;  
RT "Upregulation of a histone-like protein in dormant Mycobacterium smegmatis";  
RL Mol. Gen. Genet. 260:475-479(1998).  
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF068138; AAD13809.1; -;  
DR HSSP; P02346; 1HUU.  
DR InterPro; IPR000119; Bac\_DNABind.  
DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
DR ProDom; PD000945; Bac\_DNABind; 1.  
DR SMART; SM00411; BHL; 1.  
DR PROSITE; PS00045; HISTONE LIKE; 1.  
KW DNA-binding; DNA condensation; Repeat.  
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.  
SQ SEQUENCE 208 AA; 21230 MW; CASF577F61F7EF09 CRC64;

Query Match 43.1%; Score 71.5; DB 1; Length 208;  
Best Local Similarity 58.3%; Pred. No. 0.64;  
Matches 21; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 AKKYAKKKAAYKKEA-KAKAAEAAAEAAEAAEAA 35  
DB 111 AKKAACKPAKKAACKPAKKAACKPAKKAACKPAKKAACKPA 146

RESULT 6  
H1B STRPU STANDARD; PRT; 211 AA.  
ID H1B STRPU  
AC P15869;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Histone H1-beta, late embryonic.

Complete proteome. 8 CYTOPLASMIC (POTENTIAL).  
DOMAIN 1 29 POTENTIAL.  
TRANSMEM 9 29 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 30 372 V -> A (IN STRAIN 1479).  
FT VARIANT 48 48 K -> R (IN STRAIN 1479).  
FT VARIANT 142 142 A -> P (IN STRAIN 1479).  
FT VARIANT 165 165 A -> R (IN STRAIN 1479).  
FT VARIANT 190 190 D -> A (IN STRAIN 1479).  
FT VARIANT 203 203 D -> A (IN STRAIN 1479).  
FT VARIANT 227 227 A -> A (IN STRAIN 1479).  
FT VARIANT 232 232 A -> A (IN STRAIN 1479).  
FT VARIANT 234 234 L -> F (IN STRAIN 1479).  
FT VARIANT 249 249 I -> V (IN STRAIN 1479).  
FT VARIANT 254 254 N -> S (IN STRAIN 1479).  
FT VARIANT 306 306 T -> A (IN STRAIN 1479).  
FT VARIANT 323 323 S -> P (IN STRAIN 1479).  
FT VARIANT 333 333  
SQ SEQUENCE 372 AA; 39831 MW; 266ECP05C6C95544 CRC64;

Query Match 44.6%; Score 74; DB 1; Length 372;  
Best Local Similarity 58.1%; Pred. No. 0.58;  
Matches 18; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 5 AKKEAKKAYKKEAKAKAAEAAAEAAEAAEAA 35  
DB 158 AKRLAAAKQAEAEAKAKAEIAAQAQKEA 188

RESULT 4  
H1 LYTPI STANDARD; PRT; 210 AA.  
ID AC P06144;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Late histone H1.  
OS Lytechinus pictus (Painted sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Temnopneuroidea; Toxopneustidae;  
OC Lytechinus.  
OX NCBI\_TaxID=7653;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=87040778; PubMed=3022245;  
RA Knowles J.A., Childs G.J.;  
RT "Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus and Strongylocentrotus purpuratus";  
RL Nucleic Acids Res. 14:8121-8133(1986).  
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X04488; CAA28177.1; -;  
DR PIR; A25550; A25550.  
DR HSSP; P02259; 1HST.  
DR InterPro; IPR001386; Histone\_H1/H5.  
DR InterPro; IPR003216; Linkerhist\_N.  
DR Pfam; PF00538; linker histone; 1.  
DR ProDom; PD000373; Linkerhist\_N; 1.  
DR SMART; SM00526; H15; 1.  
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.  
SQ SEQUENCE 210 AA; 21746 MW; 08C38F6494007E2 CRC64;

Query Match 43.4%; Score 72; DB 1; Length 210;  
Best Local Similarity 59.0%; Pred. No. 0.57;  
Matches 23; Conservative 3; Mismatches 9; Indels 4; Gaps 2;

QY 1 AKKYAKK--EKAACK-AVKEAKAKAAEAARAEAYEA 35  
DB 157 AKKAARKPAKKAAKPAKKPAKKAAKPAKKAAKPA 195

RESULT 5  
DBH MYCSM STANDARD; PRT; 208 AA.  
ID DBH MYCSM  
AC Q9ZHC5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).  
GN HUP OR HLP.  
OS Mycobacterium smegmatis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1772;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 700084 / mc(2)155;  
RX MEDLINE=91110209; PubMed=9894918;  
RA Lee B.H., Murugasu-Oei B., Dick T.;  
RT "Upregulation of a histone-like protein in dormant Mycobacterium smegmatis";  
RL Mol. Gen. Genet. 260:475-479(1998).  
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AF068138; AAD13809.1; --  
DR HSSP; P02346; 1HUU.  
DR InterPro; IPR000119; Bac\_DNABind.  
DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
DR ProDom; PD000945; Bac\_DNABind; 1.  
DR SMART; SM00411; BHL; 1.  
DR PROSITE; PS00045; HISTONE LIKE; 1.  
KW DNA-binding; DNA condensation; Repeat.  
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.  
SQ SEQUENCE 208 AA; 21230 MW; CASF577F61F7EF09 CRC64;

Query Match 43.1%; Score 71.5; DB 1; Length 208;  
Best Local Similarity 58.3%; Pred. No. 0.64;  
Matches 21; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 AKKYAKKKAAYKKEA-KAKAAEAARAEAYEA 35  
DB 111 AKKAARKPAKKAAKKTATKAAKPAKKAAATKA 146

RESULT 6  
IDB STRPU  
H1B STRPU  
AC P15869;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Histone H1-beta, late embryonic.

OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
OC Strongylocentrotus.  
OX NCBI\_TaxID=7668;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88246461; PubMed=2837660;  
R Lai Z.-C., Childs G.;  
RA "Characterization of the structure and transcriptional patterns of  
RT the gene encoding the late histone subtype H1-beta of the sea urchin  
RT Strongylocentrotus purpuratus."  
RL Mol. Cell. Biol. 8:1842-1844(1988).  
CC -I- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF  
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.  
CC -I- SUBCELLULAR LOCATION: Nuclear.  
CC -I- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M20314; AAA30052.1; -.  
DR PIR; A28100; A28100.  
DR HSP; P02259; IHST.  
DR InterPro; IPR001386; Histone H1/H5.  
DR InterPro; IPR003216; Linker histone N.  
DR Pfam; PF00538; linker histone; 1.  
DR ProDom; PD000373; Linkerhist\_N; 1.  
DR SMART; SM00526; H15; 1.  
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.  
SQ SEQUENCE 211 AA; 22169 MW; 9F214581334BBE7A CRC64;  
  
Query Match 43.1%; Score 71.5; DB 1; Length 211;  
Best Local Similarity 66.7%; Pred. No. 0.65;  
Matches 20; Conservative 2; Mismatches 7; Indels 1; Gaps 1;  
  
QY 1 AKYAKKEKAAYKKEAKAKAAAEAAAKE 30  
DB 183 AKAAAK-PAAKAAKVKAAKPAKAAK 211  
  
RESULT 7  
YL17\_CABEL STANDARD; PRT; 1130 AA.  
AC Q11102;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 131.5 kDa protein C02F12.7 in chromosome X.  
GN C02F12.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Miller N.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: WEAK, TO MYOSINS.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; U41545; AAK39135.1; -.  
DR WormPep; C02F12.7; C03901.  
KW Hypothetical protein; Coiled coil.  
FT DOMAIN 121 779 COILED COIL (POTENTIAL).  
FT DOMAIN 805 1061 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1130 AA; 131485 MW; B0FD2FE3D99FB09 CRC64;  
  
Query Match 42.2%; Score 70; DB 1; Length 1130;  
Best Local Similarity 51.4%; Pred. No. 3.7;  
Matches 18; Conservative 6; Mismatches 5; Indels 6; Gaps 1;  
  
QY 6 KKEAKAKAYKKEAKAKA-----AEAAAKEAAAYE 34  
DB 1016 KKEASEKRTKEAKKEAKKEAKKEAKKEAAKE 1050  
  
RESULT 8  
TOLA\_ECOLI STANDARD; PRT; 421 AA.  
ID TOLA\_ECOLI  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tola protein.  
OS TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JM105;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levensgood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
RT their products, components of a multistep translocation system in  
RT Escherichia coli."  
RL J. Bacteriol. 171:6600-6609(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamanoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map."  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levensgood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "TolA: a membrane protein involved in colicin uptake contains an  
RT extended helical region."  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
RN [5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

RA Llobes R.;  
 RT "TolA central domain interacts with Escherichia coli porins.";  
 RL ENBO J. 15:6408-6415 (1996).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.  
 RX MEDLINE=9932679; PubMed=10404600;  
 RA Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;  
 RT "Filamentous phage infection: crystal structure of gsp in complex  
 with its coreceptor, the C-terminal domain of TolA.";  
 RL Structure 7:711-722 (1999).  
 CC -|- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A  
 COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE  
 COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL  
 BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION  
 OF BACTERIOPHAGE DNA.  
 CC -|- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE  
 AND LAMB.  
 CC -|- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M28232; AAA24683.1; -;  
 DR EMBL; AE000177; AAC73833.1; -;  
 DR EMBL; D90713; BAA35405.1; -;  
 DR PIR; J00057; J00057.  
 DR PDB; 1TOL; 20-MAY-99.  
 DR EcoGene; EG11007; tolA.  
 KW Transport; Protein transport; Bacteriocin transport; Transmembrane;  
 KW Repeat; Inner membrane; 3D-structure; Complete proteome.  
 FT DOMAIN 1 13  
 FT TRANSMEM 14 34  
 FT DOMAIN 35 421  
 FT DOMAIN 48 310  
 FT DOMAIN 311 421  
 FT DOMAIN 224 278  
 FT A(2,4).  
 SQ SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;  
 Query Match 41.9%; Score 69.5; DB 1; Length 421;  
 Best Local Similarity 52.6%; Pred. No. 1.8;  
 Matches 20; Conservative 4; Mismatches 9; Indels 5; Gaps 1;  
 QY 3 KYAKKEAKAKA-----YKKEAKAKAAFAAAKAAEAAEAA 35  
 Db 130 KQQAEEAAKAAADAKAKAEADAKAAEAAKAAADA 167  
 RESULT 9  
 HIC2\_TRYCR STANDARD; PRT; 87 AA.  
 ID HIC6\_TRYCR  
 AC P40269; P40271;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Histone H1.C6/H1.C9.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tulahuén 2;  
 RX MEDLINE=95059220; PubMed=7969272;  
 RA Aaslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,  
 RA Galanti N., Pettersson U.;  
 RT "A gene family encoding heterogeneous histone H1 proteins in  
 Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 65:317-330 (1994).  
 Query Match 41.9%; Score 69.5; DB 1; Length 421;  
 Best Local Similarity 52.6%; Pred. No. 1.8;  
 Matches 20; Conservative 4; Mismatches 9; Indels 5; Gaps 1;  
 QY 3 KYAKKEAKAKA-----YKKEAKAKAAFAAAKAAEAAEAA 35  
 Db 130 KQQAEEAAKAAADAKAKAEADAKAAEAAKAAADA 167

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L27120; AAA66479.1; -;  
 DR EMBL; L27117; AAA66481.1; -;  
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.  
 SQ SEQUENCE 87 AA; 9005 MW; 843FF90F5F73922D CRC64;  
 Query Match 41.3%; Score 68.5; DB 1; Length 87;  
 Best Local Similarity 52.6%; Pred. No. 0.62;  
 Matches 20; Conservative 3; Mismatches 12; Indels 3; Gaps 1;  
 QY 1 AKKYA---KKEKAAYKKEAKAKAAEAAKAAEAA 35  
 Db 47 AKKRAAPKKPAAAKAVTKSAKKHAAKAPKAVKKA 84  
 RESULT 10  
 HIC2\_TRYCR STANDARD; PRT; 97 AA.  
 ID HIC2\_TRYCR  
 AC P40268;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Histone H1.C2.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tulahuén 2;  
 RX MEDLINE=95059220; PubMed=7969272;  
 RA Aaslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,  
 RA Galanti N., Pettersson U.;  
 RT "A gene family encoding heterogeneous histone H1 proteins in  
 Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 65:317-330 (1994).  
 Query Match 41.3%; Score 68.5; DB 1; Length 97;  
 Best Local Similarity 52.6%; Pred. No. 0.68;  
 Matches 20; Conservative 3; Mismatches 12; Indels 3; Gaps 1;  
 QY 1 AKKYA---KKEKAAYKKEAKAKAAEAAKAAEAA 35  
 Db 57 AKKRAAPKKPAAAKAVTKSAKKHAAKAPKAVKKA 94  
 RESULT 11  
 ASR\_ENTCL STANDARD; PRT; 101 AA.  
 ID ASR\_ENTCL  
 AC O93MH6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Acid shock protein precursor.



Query Match 39.5%; Score 65.5; DB 1; Length 214;  
Best Local Similarity 52.4%; Pred. No. 2.6;  
Matches 22; Conservative 3; Mismatches 10; Indels 7; Gaps 2;

Qy 1 AKKYAKI-----EKAACKYKKEAKAKAA-EAAAKEAAAYEA 35  
||| ||| ||| ||| ||| :|| :|  
Db 111 AKKVAKKAPAKKATKAKAKKAATKAPAPKAAATKAPAKKATKA 152

Search completed: March 10, 2003, 12:17:07  
Job time : 8.97468 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 18.5338 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989a-1

Perfect score: 166

Sequence: 1 AKYAKKEKAAYKAYKKEAKAAEAKEAYEA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 21: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriophage: \*  
17: sp\_archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	81	48.8	372	2 Q9WWX1	Q9WWX1 pseudomonas
2	80.5	48.5	461	2 O68124	O68124 rhodobacter
3	79.5	47.9	98	5 Q8WQ44	Q8WQ44 leishmania
4	79.5	47.9	111	5 Q8T9R3	Q8T9R3 leishmania
5	77	46.4	275	5 O01395	O01395 drosophila
6	77	46.4	395	2 Q937K4	Q937K4 erwinia chr
7	75.5	45.5	101	5 Q9BMV8	Q9BMV8 leishmania
8	75	45.2	244	16 Q9AJX2	Q9AJX2 streptomyces
9	70.5	42.5	445	3 Q9P3Q8	Q9P3Q8 neurospora
10	70	42.2	277	16 Q9XAQ3	Q9XAQ3 streptomyces
11	70	42.2	1015	5 Q94705	Q94705 physarum po
12	69.5	41.9	78	5 Q95UN5	Q95UN5 trypanosoma
13	69	41.6	741	5 Q9X207	Q9X207 drosophila
14	68.5	41.3	83	5 Q95UN6	Q95UN6 trypanosoma
15	68.5	41.3	112	5 Q9XYI5	Q9XYI5 leishmania
16	68.5	41.3	629	2 Q9KJ98	Q9KJ98 escherichia

17	68	41.0	241	5	Q17321	Q17321 chironomus
18	68	41.0	467	2	Q9AGV7	Q9AGV7 corynebacte
19	68	41.0	1092	16	Q98FB4	Q98FB4 rhizobium l
20	67.5	40.7	81	5	Q9NFP6	Q9NFP6 trypanosoma
21	67	40.4	394	16	Q8X965	Q8X965 escherichia
22	67	40.4	515	5	Q9VBL3	Q9VBL3 drosophila
23	67	40.4	1021	16	Q9RUL0	Q9RUL0 deinococcus
24	66.5	40.1	71	5	Q9NFPJ8	Q9NFPJ8 trypanosoma
25	66.5	40.1	76	5	Q9NFPJ8	Q9NFPJ8 trypanosoma
26	66.5	40.1	76	5	Q9NFPJ4	Q9NFPJ4 trypanosoma
27	66.5	40.1	81	5	Q9NGL9	Q9NGL9 trypanosoma
28	66.5	40.1	91	5	Q9NFPJ9	Q9NFPJ9 trypanosoma
29	66.5	40.1	582	4	O60415	O60415 homo sapien
30	66.5	40.1	635	4	O60414	O60414 homo sapien
31	66.5	40.1	969	5	Q9NDI9	Q9NDI9 plasmodium
32	66.5	40.1	1052	5	Q26938	Q26938 trypanosoma
33	66	39.8	224	13	Q90ZD7	Q90ZD7 bufo bufo g
34	66	39.8	232	10	Q39576	Q39576 chlamydomon
35	66	39.8	300	10	Q8VWK4	Q8VWK4 arabidopsis
36	66	39.8	318	10	Q9LPL9	Q9LPL9 arabidopsis
37	65.5	39.5	67	16	Q8XW13	Q8XW13 raietonia s
38	65.5	39.5	481	5	Q9VWK3	Q9VWK3 drosophila
39	65.5	39.5	612	10	Q9XHE2	Q9XHE2 chlamydomon
40	65	39.2	101	2	Q9X342	Q9X342 bacillus an
41	65	39.2	322	5	Q9VPH1	Q9VPH1 drosophila
42	65	39.2	497	2	Q45302	Q45302 corynebacte
43	65	39.2	498	2	Q9ANT7	Q9ANT7 brevivbacte
44	65	39.2	901	5	Q8SUW1	Q8SUW1 encephalito
45	64.5	38.9	191	5	O46142	O46142 mytilus edu

#### ALIGNMENTS

#### RESULT 1

ID	Q9WWX1	PRELIMINARY;	PRT;	372 AA.
AC	Q9WWX1;			
DT	01-NOV-1999 (TRENBLrel. 12, Created)			
DT	01-NOV-1999 (TRENBLrel. 12, Last sequence update)			
DT	01-MAR-2002 (TRENBLrel. 20, Last annotation update)			
DE	TOLA protein.			
GN	TOLA.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=303;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RX	MEDLINE=96198174; PubMed=8626299;			
RA	Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.J.;			
RT	"The Pseudomonas putida peptidoglycan-associated outer membrane			
RT	lipoprotein (PAL) is involved in maintenance of the integrity of the			
RT	cell envelope."			
RL	J. Bacteriol. 178:1699-1706(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RA	Ramos-Gonzalez I.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RA	Rodriguez-Herva J.J.;			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RX	MEDLINE=96422022; PubMed=8824639;			
RA	Rodriguez-Herva J.J., Ramos J.J.;			
RT	"Characterization of an OprL null mutant of Pseudomonas putida."			
RL	J. Bacteriol. 178:5836-5840(1996).			

```

DR EMBL; X74218; CAB50780.1; -.
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PR00624; HISTONEHS.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match
Best Local Similarity 56.1%; Pred. No. 0.21;
Matches 23; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

Qy 1 AKKYA-----KKEAKAKYKKEAKAKAAEAAAEAAAEAA 35
Db 160 AKKADEAKKAAEEAAKAAAEAAEAKKAAAEADAKKAAEEA 200

RESULT 2
O68124 PRELIMINARY; PRT; 461 AA.
AC O68124;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Hypothetical 49.5 kDa protein.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003;
RX MEDLINE=97404404; PubMed=9256491;
RA Vilek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter
RL capsulatus SB1003.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388 (1997).
DR EMBL; AF010496; AAC16214.1; -.
KW Hypothetical protein.
SQ SEQUENCE 461 AA; 49516 MW; DD8DA03418BC0368 CRC64;

Query Match
Best Local Similarity 60.0%; Pred. No. 0.29;
Matches 21; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

Qy 1 AKKYAKKEAKAKYKKEAKAKAAEAAAEAAAEAA 35
Db 30 AKEWVEKEAAAKDAEAAEAKAKAR-EAAAKDAARA 63

RESULT 3
Q8WQ44 PRELIMINARY; PRT; 98 AA.
AC Q8WQ44;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE 18 kDa nuclear protein.
GN LNP18.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RA Tzortzakos N., Papageorgiou F.T., Tzinia A.K., Soteriadou K.P.;
RT "Identification and characterization of a novel Leishmania gene
RL encoding for a putative histone H1-like transcription factor.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237814; CAD21431.1; -.
KW Nuclear protein.
SQ SEQUENCE 98 AA; 9999 MW; 0A4AB93089D6C261 CRC64;

Query Match
Best Local Similarity 47.9%; Score 79.5; DB 5; Length 98;
Matches 23; Conservative 3; Mismatches 5; Indels 3; Gaps 2;

```

```

Qy 1 AKKYAKK--EKAACKYKKEAKAKAAEAAAEAA 32
Db 63 AKKVAKKPAKKAACKKPAKPAK-KAACKAAKAA 95

RESULT 4
Q8T9R3 PRELIMINARY; PRT; 111 AA.
AC Q8T9R3;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE 18 kDa nuclear protein.
OS Leishmania infantum.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RA Papageorgiou F., Soteriadou K.;
RT "Identification of a Leishmania infantum gene encoding for an histone
RL H1-like nuclear protein.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF469106; AAL76335.1; -.
KW Nuclear protein.
SQ SEQUENCE 111 AA; 11162 MW; 16168F3B54960E83 CRC64;

Query Match
Best Local Similarity 47.9%; Score 79.5; DB 5; Length 111;
Matches 23; Conservative 3; Mismatches 5; Indels 3; Gaps 2;

Qy 1 AKKYAKK--EKAACKYKKEAKAKAAEAAAEAA 32
Db 76 AKKVAKKPAKKAACKKPAKPAK-KAACKAAKAA 108

RESULT 5
O01395 PRELIMINARY; PRT; 275 AA.
AC O01395;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE Axoneme-associated protein MST101(3).
GN MST101(3) OR DHMST101.
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A.
RA Neesen J., Heinlein U.A.O., Buenemann H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR EMBL; U85627; AAB51369.1; -.
DR FlyBase; FBgn020732; Dhyd\mst101(3).
FT Spm; Repeat; Multigene family.
FT DOMAIN 64 255
FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-
FT [AE]-X
SQ SEQUENCE 275 AA; 30436 MW; 76BA7B2A2DF732C CRC64;

Query Match
Best Local Similarity 46.4%; Score 77; DB 5; Length 275;
Matches 20; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 AKKYAKKPAKKAACKKPAKPAK-KAACKAAKAA 35
Db 196 AKKAAEEAAKKAAEVAACKKAAEAAKKAEEA 230

```





Search completed: March 10, 2003, 12:25:52  
Job time : 20.5338 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 18.2384 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-1  
Perfect score: 166  
Sequence: 1 AKYAKKEKAANKYKKEAKAKAAAEAAAEAYEA 35

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq 101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166	100.0	35	21	Copolymer molecule
2	108	65.1	21	AAV82571	Copolymer molecule
3	108	65.1	86	21	Copolymer molecule
4	105.5	63.6	66	21	Copolymer molecule
5	105	63.3	109	21	Copolymer molecule
6	104.5	63.0	56	21	Copolymer molecule
7	104.5	63.0	77	21	Copolymer molecule
8	86.5	52.1	427	22	C albicans apoptos
9	76	45.8	154	11	Recombinant copoly
10	73.5	44.3	40	21	Peptide modulating

11	71	42.8	106	11	AA06446
12	69.5	41.9	146	22	AA02508
13	69.5	41.9	165	22	AA09197
14	69	41.6	741	22	AB06208
15	67.5	40.7	32	16	AA09180
16	67.5	40.7	32	18	AA06686
17	67.5	40.7	32	21	AA08168
18	67.5	40.7	33	16	AA090181
19	67.5	40.7	33	18	AA06688
20	67	40.4	515	22	AB061894
21	66.5	40.1	582	23	AA084310
22	66	39.8	242	21	AA037166
23	66	39.8	242	21	AA038490
24	66	39.8	245	21	AA037165
25	66	39.8	245	21	AA038489
26	66	39.8	300	21	AA037164
27	66	39.8	300	21	AA038488
28	65.5	39.5	214	20	AA034055
29	65.5	39.5	214	21	AA057353
30	65.5	39.5	481	22	AB063417
31	65	39.2	223	20	AA014928
32	65	39.2	322	22	AB059960
33	65	39.2	498	22	AA091846
34	64.5	38.9	203	18	AA014557
35	64.5	38.9	205	21	AA020575
36	64.5	38.9	769	22	AA034982
37	64.5	38.9	782	22	AA033397
38	63.5	38.3	621	20	AA07029
39	63	38.0	607	22	AB062173
40	63	38.0	2451	22	AB071574
41	62.5	37.7	29	16	AA090176
42	62.5	37.7	29	16	AA090178
43	62.5	37.7	29	18	AA06687
44	62.5	37.7	29	18	AA06697
45	62.5	37.7	262	17	AA06913

ALIGNMENTS

RESULT 1

AAV82571 ID AAV82571 standard; peptide; 35 AA.

XX AAV82571;

XX 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:1.

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
XX glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
XX antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
XX antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
XX pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX WO200018794-A1.

XX 06-APR-2000.

XX 24-SEP-1999; 99WO-US22402.

XX 25-SEP-1998; 98US-0101693.

XX (YEDA ) YEDA RES & DEV CO LTD.

XX (TEVA-) TEVA PHARM USA INC.







XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
OS Unidentified.  
PN WO200018794-A1.  
XX  
PD 06-APR-2000.  
XX  
PF 24-SEP-1999; 99WO-US22402.  
XX  
PR 25-SEP-1998; 98US-0101693.  
XX  
PA (YEDA ) YEDA RES & DEV CO LTD.  
PA (TEVA-) TEVA PHARM USA INC.  
XX  
PI Gad A, Lis D;  
XX  
DR WPI; 2000-317499/27.  
XX  
PT Copolymer 1 related polypeptides used as molecular weight markers for  
PT glatiramer acetate and for treatment and prevention of immune diseases  
XX  
PS Claim 10; Page 14; 72pp; English.  
XX  
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular  
CC weight of a copolymer (CP), which has an identified molecular weight  
CC and an amino acid composition corresponding to the copolymer. The  
CC polypeptides of the invention are used as molecular weight markers for  
CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
CC used for treating and preventing immune diseases in a mammal. Autoimmune  
CC diseases which may be treated include either cell-mediated or  
CC antibody-mediated diseases. Such diseases include arthritic conditions,  
CC demyelinating diseases and inflammatory conditions, e.g. multiple  
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
CC lupus erythematosus. Mediated-mediated diseases which can be treated  
CC include host-versus-graft disease, graft-versus-host disease, and  
CC delayed-type hypersensitivity. The polypeptides of the invention have  
CC defined molecular weights and physical properties which are analogous to  
CC glatiramer acetate molecules, which makes them ideal for use as  
CC molecular weight markers.  
XX  
SQ Sequence 109 AA;  
Query Match 63.3%; Score 105; DB 21; Length 109;  
Best Local Similarity 62.2%; Pred. No. 3.6e-05;  
Matches 28; Conservative 2; Mismatches 5; Indels 10; Gaps 2;  
QY 1 AKKYAKKEAAKAY-----KKEA-----KAKAAEAKEAAEAYEA 35  
Db 65 AKKYAKAAKAEKGYAAAEAAKAEAAKAYAAKAAKAAKAAEAYEA 109  
RESULT 6  
AAY82573  
ID AAY82573 standard; peptide; 56 AA.  
XX

AC AAY82573;  
XX  
DT 28-JUL-2000 (first entry)  
XX  
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.  
XX  
KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX  
OS Unidentified.  
XX  
PN WO200018794-A1.  
XX  
PD 06-APR-2000.  
XX  
PF 24-SEP-1999; 99WO-US22402.  
XX  
PR 25-SEP-1998; 98US-0101693.  
XX  
PA (YEDA ) YEDA RES & DEV CO LTD.  
PA (TEVA-) TEVA PHARM USA INC.  
XX  
PI Gad A, Lis D;  
XX  
DR WPI; 2000-317499/27.  
XX  
PT Copolymer 1 related polypeptides used as molecular weight markers for  
PT glatiramer acetate and for treatment and prevention of immune diseases  
XX  
PS Claim 10; Page 14; 72pp; English.  
XX  
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular  
CC weight of a copolymer (CP), which has an identified molecular weight  
CC and an amino acid composition corresponding to the copolymer. The  
CC polypeptides of the invention are used as molecular weight markers for  
CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
CC used for treating and preventing immune diseases in a mammal. Autoimmune  
CC diseases which may be treated include either cell-mediated or  
CC antibody-mediated diseases. Such diseases include arthritic conditions,  
CC demyelinating diseases and inflammatory conditions, e.g. multiple  
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
CC lupus erythematosus. Mediated-mediated diseases which can be treated  
CC include host-versus-graft disease, graft-versus-host disease, and  
CC delayed-type hypersensitivity. The polypeptides of the invention have  
CC defined molecular weights and physical properties which are analogous to  
CC glatiramer acetate molecules, which makes them ideal for use as  
CC molecular weight markers.  
XX  
SQ Sequence 56 AA;  
Query Match 63.0%; Score 104.5; DB 21; Length 56;  
Best Local Similarity 51.8%; Pred. No. 2.1e-05;  
Matches 29; Conservative 4; Mismatches 2; Indels 21; Gaps 2;  
QY 1 AKKYAKKEKA---AKKAYKK-----EAKAAEAKEAAEAYEA 35  
Db 1 AKKYAKKEKAYAKAEKAAKAEAAKAEAAKAEAAKAEAAEAYEA 56

RESULT 7  
AA082575  
ID AA082575 standard; peptide; 77 AA.  
XX  
AC AA082575;  
XX  
DT 28-JUL-2000 (first entry)  
XX  
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.  
XX  
KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX  
OS Unidentified.  
XX  
XX WO200018794-A1.  
XX  
XX 06-APR-2000.  
XX  
XX 24-SEP-1999; 99WO-US22402.  
XX  
XX 25-SEP-1998; 98US-0101693.  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX (TEVA-) TEVA PHARM USA INC.  
XX  
XX Gad A, Lis D;  
XX  
XX WPI; 2000-317499/27.  
XX  
XX Copolymer 1 related polypeptides used as molecular weight markers for  
XX glatiramer acetate and for treatment and prevention of immune diseases  
XX  
XX Claim 10; Page 14; 72pp; English.  
XX  
XX AA082571 to AA082577 represent specifically claimed copolymer molecular  
XX weight TV-marker polypeptides from the present invention. The present  
XX invention describes polypeptides (I) for determining the molecular  
XX weight of a copolymer (CP), which has an identified molecular weight  
XX and an amino acid composition corresponding to the copolymer. The  
XX polypeptides of the invention are used as molecular weight markers for  
XX glatiramer acetate related tetrapolymers. The polypeptides may also be  
XX used for treating and preventing immune diseases in a mammal. Autoimmune  
XX diseases which may be treated include either cell-mediated or  
XX antibody-mediated diseases. Such diseases include arthritic conditions,  
XX demyelinating diseases and inflammatory conditions, e.g. multiple  
XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
XX anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
XX uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
XX lupus erythematosus. Mediated-mediated diseases which can be treated  
XX include host-versus-graft disease, graft-versus-host disease, and  
XX delayed-type hypersensitivity. The polypeptides of the invention have  
XX defined molecular weights and physical properties which are analogous to  
XX glatiramer acetate molecules, which makes them ideal for use as  
XX molecular weight markers.  
XX  
XX Sequence 77 AA;

Query Match 63.0%; Score 104.5; DB 21; Length 77;  
Best Local Similarity 65.8%; Pred. No. 2.9e-05;  
Matches 25; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

QY 1 AKKYAKKEKAKKAY---KKEAKAKAAEAEEAAKAAEAA 35  
| | | | | | | | | | : : : : : : : : : : : : : : : :  
Db 40 AKKYAKAAKEKKEYAAAAEAKYKAAEAEEAAKAAEAA 77  
| | | | | | | | | | : : : : : : : : : : : : : : : :  
RESULT 8  
AAG70868  
ID AAG70868 standard; Protein; 427 AA.  
XX  
AC AAG70868;  
XX  
DT 27-JUL-2001 (first entry)  
XX  
DE C albicans apoptosis associated protein #48.  
XX  
KW Yeast; fungus; apoptosis; infection; proliferative disease;  
KW vaccine; autoimmune disease; ischaemia; neurodegeneration.  
XX  
OS Candida albicans.  
XX  
XX WO200102550-A2.  
XX  
XX 11-JAN-2001.  
XX  
XX 03-JUL-2000; 2000WO-BE00077.  
XX  
XX 01-JUL-1999; 99EP-0870141.  
XX  
XX (JANC ) JANSSEN PHARM NV.  
XX  
XX Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;  
XX Nelissen BJM, Reekmans RJ;  
XX WPI; 2001-367042/38.  
XX N-PSDB; AAH29904.  
XX  
XX Yeast and fungal nucleic acids encoding proteins involved in a pathway  
XX leading to programmed cell death, useful for treating proliferative  
XX disorders, yeast and fungal infections, or for preventing apoptosis in  
XX certain diseases -  
XX  
XX Claim 24; Fig 2; 218pp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of apoptosis associated proteins from the yeast *Saccharomyces*  
XX *cerevisiae* and the fungus *Candida albicans*. These can be used to identify  
XX treatments for fungal and yeast infections, for proliferative diseases  
XX and for apoptosis related diseases such as autoimmune diseases, ischaemia  
XX and neurodegeneration. The present sequence is one of the *C. albicans*  
XX proteins of the invention.  
XX  
XX Sequence 427 AA;  
XX  
XX Query Match 52.1%; Score 86.5; DB 22; Length 427;  
XX Best Local Similarity 65.7%; Pred. No. 0.02;  
XX Matches 23; Conservative 3; Mismatches 8; Indels 1; Gaps 1;  
QY 2 KKYAKKEKAKKAYKKEAKAKAA-EAAKAAEAA 35  
| | | | | | | | | | : : : : : : : : : : : : : : : :  
Db 283 KEKAKKEKAAKWEKSGSRRAAEAAKAAEAA 317  
| | | | | | | | | | : : : : : : : : : : : : : : : :  
RESULT 9  
AAR06445  
ID AAR06445 standard; protein; 154 AA.  
XX  
AC AAR06445;  
XX  
XX 03-JAN-1991 (first entry)  
XX  
XX Recombinant copolymer 1-77, myelin basic protein analogue.  
XX



```
XX PI Cook KS;
XX PF WPI; 1990-255848/34.
XX DR N-PSDB; AAQ06446.
XX PR Producing genes encoding random polymers of aminoacid(s) - for
XX PT producing recombinant polypeptide(s) with biological and/or
XX PT immunological activity
XX PS Disclosure; Fig 12; 25pp; English.
XX CC To improve the expression of rCOP-1 polypeptides in E. coli, genes
XX CC coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2delTan
XX CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to
XX CC express Protein A. The resulting plasmids encode fusion proteins
XX CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.
XX CC A methionine residue occurs between the Protein A and rCOP-1
XX CC sequences, originating from the 5' linker sequence, in order that
XX CC the COP-1 polypeptide may be cleaved from the fusion protein.
XX CC rCOP-1-19 contains oligonucleotide duplexes encoding the following
XX CC segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA KEA, and AAA. The
XX CC N-terminal alanine residue is left behind following CNBr cleavage of the
XX CC fusion protein.
XX CC The product prevents or arrests experimental autoimmune
XX CC encephalomyelitis. They are used to prevent, arrest or control a
XX CC demyelinating disorder, e.g. multiple sclerosis. They may also
XX CC be used as additives to hair care products to confer beneficial
XX CC effects on damaged hair or as supplements for diets deficient in
XX CC certain amino acids.
XX CC See also AAQ05664.
XX SQ Sequence 106 AA;
Query Match 42.8%; Score 71; DB 11; Length 106;
Best Local Similarity 48.9%; Pred. No. 0.32; Mismatches 5; Indels 12; Gaps 2;
Matches 22; Conservative 5;
QY 3 KYAKKEKAAYKKE-AKA-----KAAFAAAKEAAEYEA 35
| : ||||| : |||
Db 2 KAAEKAAYKAYEAKAKAYEAKKAEKAEKAEKAEKAEKAEKAE 46
| : ||||| : |||

RESULT 12
AAM25508
ID AAM25508 standard; Protein; 146 AA.
AC AAM25508;
XX 16-OCT-2001 (first entry)
XX Human protein sequence SEQ ID NO:1023.
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX antibacterial; endocrine; cardiant; central nervous system; virucide;
XX anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
XX antiaggregant; haemostatic; vulnerary; antilucer; osteopathic; eczema;
XX dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX thrombocytopenia; osteoporosis; severe combined immunodeficiency;
XX allergic rhinitis; diabetes; multiple sclerosis; depression;
XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX neurological disorder.
XX Homo sapiens.
XX OS.
XX WO200153455-A2.
XX PN
XX

PD 26-JUL-2001.
XX 22-DEC-2000; 2000WO-US35017.
XX 23-DEC-1999; 99US-0471275.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457603/49.
XX N-PSDB; AAH99449.
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 20; Page 210; 1217pp; English.
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX AAM25963. The proteins can have activities based on the tissues and
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
XX antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis; viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX neurological disorders.
XX SQ Sequence 146 AA;
Query Match 41.9%; Score 69.5; DB 22; Length 146;
Best Local Similarity 52.6%; Pred. No. 0.65;
Matches 20; Conservative 4; Mismatches 9; Indels 5; Gaps 1;
QY 3 KYAKKEKAAYKKE-----YKKEAKAKAAFAAAKEAAEYEA 35
| : ||||| : |||
Db 107 KQKQAEAAKAAADAKAKAEADAKAEAAKAAKAAADA 144
| : ||||| : |||

RESULT 13
AAG91997
ID AAG91997 standard; Protein; 165 AA.
AC AAG91997;
XX 26-SEP-2001 (first entry)
XX C glutamicum protein fragment SEQ ID NO: 5751.
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX Corynebacterium glutamicum.
XX OS
XX EPI108790-A2.
XX PN
XX 20-JUN-2001.
XX PD
XX 18-DEC-2000; 2000EP-0127698.
XX PF
```



XX Novel peptides are claimed which contain 20-40 amino acids and  
CC which have a total cationic charge of less than +21 (pref. +16 to +18;  
CC esp. +18) as determined by the number of positively charged amino acids  
CC in the sequence. Preferably the positive charges are grouped in clusters  
CC separated by neutral amino acids. Especially the peptides contain 28-32  
CC amino acids having blocks of 2-4 positively charged amino acids  
CC separated by blocks of 2-6 neutral acids. Alternatively the positive  
CC charge may be distributed evenly or randomly along the peptide sequence.  
CC In particular the peptides are analogues of n-protamine (total cationic  
CC charge = +21) in which selected arginine residues have been replaced  
CC with uncharged amino acids and other arginine residues have been  
CC replaced by other positively charged amino acid residues, preferably  
CC lysine. The peptides reverse the effect of low mol. wt. heparin (LMWH)  
CC anticoagulation and hence can be used medically to prevent bleeding  
CC after the conclusion of clinical procedures using heparin therapy. They  
CC are less toxic than n-protamine since the reduced positive charge gives  
CC an improved efficiency to toxicity ratio; and they may be more effective  
CC than n-protamine in their anti-LMWH action.  
CC The present sequence (total cationic charge = +18) is a specific example  
CC of the new polypeptides.

XX  
SQ Sequence 32 AA;

Query Match 40.7%; Score 67.5; DB 16; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.24;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 AKKYAKK-EKAACKAYKKEAKAKAEAAAKE 30  
||| ||| : ||||| | ||| | :  
Db 2 AKKAKKAKKAKKAKKAKKAKKAKKAKK 32

Search completed: March 10, 2003, 12:21:27  
Job time : 19.2384 secs

THIS PAGE BLANK (USPTO)







```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 372
; TYPE: PRT
; ORGANISM: H. influenzae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: outer membrane integrity protein (tolA)
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|1573353
US-09-820-843A-8

Query Match      44.6%; Score 74; DB 9; Length 372;
Best Local Similarity 58.1%; Pred. No. 0.12;
Matches 18; Conservative 5; Mismatches 8; Indels 0; Gaps 1;

QY   5 AKKEAKKAYKKAEAKAAEAEEAAAKEAAEYA 35
     || :|||:|||:|||:|||:||||
Db   158 AKRLAAQAQEAEAKAAEAIEAAQKAKQEA 188

RESULT 9
US-09-738-626-5751
; Sequence 5751, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5751
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5751

Query Match      41.9%; Score 69.5; DB 9; Length 165;
Best Local Similarity 50.0%; Pred. No. 0.17;
Matches 19; Conservative 6; Mismatches 8; Indels 5; Gaps 1;

QY   2 KKYAKKEKAAYKKAEAKAKAA-----EAAAKEAAEY 34
     ||::|||:|||:|||:|||:|||
Db   126 KKQAREDKAEAKAAEAKEAAAEESEAPAEAAAAE 163

RESULT 10
US-09-919-497-100
; Sequence 100, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
```





**THIS PAGE BLANK (USPTO)**



;  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
US-08-152-488-13

Query Match 40.7%; Score 67.5; DB 1; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.05;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKK-EKAACKAYKKEAKAKAAEAAAKE 30  
||| ||| :||||| | ||| ||| :  
Db 2 AKAAKAKAKAKAKAKAKAKAKAKAKAKK 32

## RESULT 2

US-08-303-025-15  
; Sequence 15, Application US/08303025  
; Patent No. 5614494  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 150 West Jefferson, Suite 2500  
; CITY: Detroit  
; STATE: Michigan  
; COUNTRY: United States of America  
; ZIP: 48226-4415

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS v.6.22  
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/303,025  
; FILING DATE: 08-SEPT-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06829  
; FILING DATE: 14-AUG-1992  
; APPLICATION NUMBER: US 08/152,488  
; FILING DATE: 12-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 313-496-7622  
; TELEFAX: 313-496-8454  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: N/A  
; PUBLICATION INFORMATION:  
; AUTHORS: N/A  
; TITLE: N/A  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
US-08-303-025-15

Query Match 40.7%; Score 67.5; DB 1; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.05;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKK-EKAACKAYKKEAKAKAAEAAAKE 30  
||| ||| :||||| | ||| ||| :  
Db 2 AKAAKAKAKAKAKAKAKAKAKAKAKAKK 32

## RESULT 3

US-08-677-304-13  
; Sequence 13, Application US/08677304  
; Patent No. 5721212  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 512 Springfield Avenue  
; CITY: Cranford  
; STATE: New Jersey  
; COUNTRY: United States of America  
; ZIP: 07016-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 6; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/677,304  
; FILING DATE:  
; CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/152,488  
; FILING DATE: 12-NOV-1993  
; APPLICATION NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REFERENCE/DOCKET NUMBER: RM-7WG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-276-3344  
; TELEFAX: 908-276-5543  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: No. 5721212 Relevant  
; TOPOLOGY: No. 5721212 Relevant  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: N/A  
; PUBLICATION INFORMATION:  
; AUTHORS: N/A  
; TITLE: N/A  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
US-08-677-304-13

Query Match 40.7%; Score 67.5; DB 1; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.05;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKK-EKAACKAYKKEAKAKAAEAAAKE 30  
||| ||| :||||| | ||| ||| :  
Db 2 AKAAKAKAKAKAKAKAKAKAKAKAKAKK 32

## RESULT 4

US-08-436-703B-2



```
; Sequence 2, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 6601 Woodward Avenue
; CITY: Suite 1525
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk 1.44Mb, 3.5"
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6;
; SOFTWARE: ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,703B
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: N/A
; FILING DATE: N/A
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: 7WK-060548-00233
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-965-1976
; TELEFAX: 313-965-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; US-08-436-703B-2

Query Match 40.7%; Score 67.5; DB 2; Length 32;
Best Local Similarity 61.3%; Pred. No. 0.05;
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 AKYAKK-EKAAYKAYKKEAKAKAAEAAAKE 30
   ||| ||| : ||| ||| ||| ||| |||
Db 2 AKAAKAKAKAAKAAKAAKAAKAAKAAK 32

RESULT 5
US-08-303-025-16
; Sequence 16, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
```

```
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-7622
; TELEFAX: 313-496-8454
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-16

Query Match 40.7%; Score 67.5; DB 1; Length 33;
Best Local Similarity 61.3%; Pred. No. 0.051;
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 AKYAKK-EKAAYKAYKKEAKAKAAEAAAKE 30
   ||| ||| : ||| ||| ||| ||| |||
Db 3 AKAAKAKAKAAKAAKAAKAAKAAKAAK 33

RESULT 6
US-08-436-703B-4
; Sequence 4, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 6601 Woodward Avenue
; CITY: Suite 1525
; STATE: Michigan
; COUNTRY: United States of America
```



```
/ ; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-095-855-201

Query Match          39.2%; Score 65; DB 4; Length 223;
Best Local Similarity 54.3%; Pred No. 0.75;
Matches 19; Conservative 4; Mismatches 10; Indels 2; Gaps 1;

QY   1 AKKYAKKEAAKKAYKKAEEAAGAAAEEAAVEA 35
DB   112 ARKAARKAPAKAAKAAKPAK--KAPAKAATKA 144
      |.:||| |:| |:| |:| |:| |:| |:| |:| :|.|.:|.:|.:|.:|.:|.:|.
RESULT 9
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-205-426-201

Query Match          39.2%; Score 65; DB 4; Length 223;
Best Local Similarity 54.3%; Pred. No. 0.75;
Matches 19; Conservative 4; Mismatches 10; Indels 2; Gaps 1;

QY   1 AKKYAKKEAAKKAYKKAEEAAGAAAEEAAVEA 35
DB   112 ARKAARKAPAKAAKAAKPAK--KAPAKAATKA 144
      |.:||| |:| |:| |:| |:| |:| |:| |:| :|.|.:|.:|.:|.:|.:|.:|.
RESULT 10
US-08-152-488-10
; Sequence 10, Application US/08152488
; Patent No. 5534619
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 512 Springfield Avenue
; CITY: Cranford
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07016-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: MS-DOS
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,488
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; OPERATING SYSTEM: MS-DOS
```

```
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: RM-7WG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-276-3344
; TELEFAX: 908-276-5543
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-152-488-11

Query Match 37.7%; Score 62.5; DB 1; Length 29;
Best Local Similarity 62.1%; Pred. No. 0.17;
Matches 18; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKKEKAAYKAYKKEAKAKAEAAK 29
   ||| ||| ||| ||| ||| ||| |||
Db 2 AKKAAKKAAYKAKKAAKAKKAAK 29

RESULT 12
US-08-303-025-10
; Sequence 10, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-8454
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
```

```
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-10

Query Match 37.7%; Score 62.5; DB 1; Length 29;
Best Local Similarity 62.1%; Pred. No. 0.17;
Matches 18; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKKEKAAYKAYKKEAKAKAEAAK 29
   ||| ||| ||| ||| ||| ||| |||
Db 2 AKKAAKKAAYKAKKAAKAKKAAK 29

RESULT 13
US-08-303-025-11
; Sequence 11, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-8454
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 13.1962 Seconds  
(without alignments) 327.825 Million cell updates/sec

Title: US-09-816-989A-2  
 Perfect score: 213  
 Sequence: 1 AKKYAKKAAEKAKKAYKAA.....AKYEKAAAEKAAKEAAVEA 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 2832224

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10
                  Listing first 45

```

```
Database : PIR_73:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	103.5	48.6	210	2	A25550	histone H1 - sea u
2	98	46.0	394	2	F90725	membrane spanning
3	98	46.0	394	2	G85576	membrane spanning
4	98	46.0	421	2	JV0057	tolA protein - Esc
5	94.5	44.4	239	2	AE1317	hypothetical prote
6	90	42.3	243	2	AE1689	hypothetical prote
7	90	42.3	1390	2	S51364	sperm tail-specific
8	89.5	42.0	211	2	A28100	histone H1-beta, e
9	89.5	42.0	214	2	G70673	probable hupB - My
10	88	41.3	899	2	B87553	DNA topoisomerase
11	86.5	40.6	384	2	B43592	outer membrane pro
12	86	40.4	199	2	F70742	hypothetical prote
13	86	40.4	344	2	S34153	mst101-1 protein -
14	86	40.4	347	2	E83525	TolA protein PA097
15	85	39.9	182	2	S61926	histone H1 homolog
16	82	38.5	248	1	HSUR1P	histone H1, gonada
17	82	38.5	262	2	T35503	1-acetylgluceryl-3-p
18	82	38.5	568	2	T39675	asparaginyl-trna s
19	81.5	38.3	311	2	T17698	hypothetical prote
20	80.5	37.8	277	2	T34625	probable NLP/P60 f
21	80.5	37.8	445	2	T50972	probable zuotin [1
22	80.5	37.8	952	2	T36664	probable DNA topoi
23	80	37.6	194	1	HSTR1	histone H1 - trout
24	80	37.6	206	1	HSTR1R	histone H1 - rainb
25	80	37.6	291	2	AB1487	hypothetical prote
26	79.5	37.3	217	2	A26721	histone H1-gamma,
27	79	37.1	62	2	T30977	hypothetical prote
28	79	37.1	205	2	S19114	cger-1 protein - C
29	79	37.1	444	2	S40436	histone H1 - midge

## ALIGNMENTS

## RESULT 1

A25550  
histone H1 - sea urchin (*Lytechinus pictus*)  
C;Species: *Lytechinus pictus* (painted urchin)  
C;date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999  
C;Accession: A25550  
R;Knowles, J.A.; Childs, G.J.  
Nucleic Acids Res. 14, 8121-8133, 1986  
A;Title: Comparison of the late H1 histone genes of the sea urchins *Lytechinus pictus* and *Paracentrotus lividus*.  
A;Reference number: A25550: MUID:87040778: PMID:3022245

Query Match 48.6%; Score 103.5; DB 2; Length 210;  
Best Local Similarity 65.1%; Pred. No. 0.0053;  
Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 AKKYAKKAEEK - AKKAYKAAEAKKAAKYKAAAEKAAAKKAA 42  
||| ||| : | |||| | |||| | ||| : ||  
Db 166 AKKAAKPAAKKPAKKAAPKAAKPAKKAAPKAAKPAKKAAPKAAK 208

## RESULT 2

F90725  
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli  
C;date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: F90725  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawara, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A;Reference number: A95629, UID:21156231, PMID:11259796  
A;Accession: F90725  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-394 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA34197.1; PID:gl3360233; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: ECs0774

Query Match 46.0%; Score 98; DB 2; Length 394;  
Best Local Similarity 65.3%; Pred. No. 0.028;  
Matches 32: Conservative 2; Mismatches 9: Indels 6: Gaps 3;

QY 2 KKYAKKAKAEKA -- -- KK -- AYKAAEAKKAAYEKAAAEKAAAEAYEA 45

Db 220 KKAEEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAKAAAEAA 267

RESULT 3

G85576

membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Nov-2001

C;Accession: G85576

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: G85576

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-394 <STO>

A;Cross-references: GB:AE005174; NID:g12513672; PIDN:AAG55075.1; GSPDB:GN00145; UMGF:209

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: tolA

Query Match 46.0%; Score 98; DB 2; Length 394;

Best Local Similarity 65.3%; Pred. No. 0.028;

Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

QY 2 KKYAKKAKAEKA---KK--AYKAAEKAAKAYEKAAAEKAAAEKAAAEAA 45

Db 220 KKAEEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAKAAAEAA 267

RESULT 4

JV0057

tolA protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002

C;Accession: JV0057; B64810

R;Levengood, S.K.; Webster, R.E.

J. Bacteriol. 171, 6600-6609, 1989

A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their products

A;Reference number: JV0057; MUID:90078104; PMID:2687247

A;Accession: JV0057

A;Molecule type: DNA

A;Residues: 1-421 <LEV>

A;Cross-references: GB:M28232; NID:g148018; PIDN:AAA24683.1; PID:g148019

A;Experimental source: strain JM105

A;Note: the authors translated the initiation codon GTG for residue 1 as Val

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64810

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <BLAT>

A;Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC73833.1; PID:g1786960;

A;Experimental source: strain K-12, substrain MG1655

C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach the cell surface

C;Genetics:

A;Gene: tolA

A;Map position: 17 min

A;Start codon: GTG

C;Keywords: nucleotide binding; P-loop; transmembrane protein

F;14-34/Domain: transmembrane #status predicted <MSS>

F;78-301/Domain: helical #status predicted <HSP>

F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 46.0%; Score 98; DB 2; Length 421;

Best Local Similarity 65.3%; Pred. No. 0.029;

Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

QY 2 KKYAKKAKAEKA---KK--AYKAAEKAAKAYEKAAAEKAAAEKAAAEAA 45

Db 247 KKAEEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAKAAAEAA 294

RESULT 5

AE1317

hypothetical protein lmo1941 [imported] - Listeria monocytogenes (strain EGD-e)

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AE1317

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AE1317

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-239 <GLA>

A;Cross-references: GB:NC\_003210; PIDN:CAD00019.1; PID:g16411394; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo1941

Query Match 44.4%; Score 94.5; DB 2; Length 239;

Best Local Similarity 56.2%; Pred. No. 0.04;

Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;

QY 2 KKYAKKAKAEKAKK---AYKAAEKK--AAKYEKAAAEKAAAEKAAAE 44

Db 124 KAAAEKAAEDKKQBEDAVKAAAKKEQEAABEKAADKAAAEKAAAE 171

RESULT 6

AE1689

hypothetical protein homolog lin2055 [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AE1689

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AE1689

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-243 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC97285.1; PID:g16414556; GSPDB:GN00178

A;Experimental source: strain Clip11262

C;Genetics:

A;Gene: lin2055

Query Match 42.3%; Score 90; DB 2; Length 243;

Best Local Similarity 45.5%; Pred. No. 0.11;

Matches 30; Conservative 3; Mismatches 11; Indels 22; Gaps 2;

QY 1 AKKYA-KYAKAEKAKYKAAAEKKA-----AKYEKAAAEKAAA 38

Db 106 AKKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAE 165

QY 39 KEAAVE 44

Db 166 DKAAKE 171



G70673  
probable hupB - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: G70673  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

Qy 1 AKKYAKKAAEKAKKAYKAAAEKAAKAYEKAAAEKAAAEKAAAEKAAAE 44  
Db 178 AKEKAAEKAAKAAKAAEKAAEKAAKAAKAAKAAEKAAEKAAAE 220

RESULT 12  
F70742  
hypothetical protein RV0475 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: F70742  
R:Cole, S.T.; Broech, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70742  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-199 <COL>  
A:Cross-references: GB:277162; GB:AL123456; NID:g3261606; PIDN:CAB00936.1; PID:e255021;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV0475

Query Match 40.4%; Score 86; DB 2; Length 199;  
Best Local Similarity 65.6%; Pred. No. 0.21;  
Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 11 EKAKKAYKAAAEKAAKAYEKAAAEKAAAEKAA 42  
Db 161 KKAAPAKKAAAPAKKAAAPAKKAAAPAKKAA 192

RESULT 13  
S34153  
met101-1 protein - fruit fly (Drosophila hydei)  
C:Species: Drosophila hydei  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S34153  
R:Neesen, J.; Heinlein, U.A.O.; Bueneemann, H.  
submitted to the EMBL Data Library, June 1993  
A:Reference number: S34153  
A:Accession: S34153  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-344 <NEE>  
A:Cross-references: EMBL:X73480; NID:g313199; PID:g313200  
C:Genetics:  
A:Gene: FlyBase:Dhyd/met101  
A:Cross-references: FlyBase:FBgn0011816  
C:Superfamily: neurofilament triplet H protein

Query Match 40.4%; Score 86; DB 2; Length 344;  
Best Local Similarity 57.8%; Pred. No. 0.33;  
Matches 26; Conservative 3; Mismatches 12; Indels 4; Gaps 2;

Qy 2 KKYAKKAK-AEKAKAYKAAAEK---AAKYKAAAEKAAAEKAA 42  
Db 117 KEAAEKKKCAEAKEKAAEKKEKCAEAKEKAAEKKEKCAEA 161

RESULT 14  
E83525  
TolA protein PA0971 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83525  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: E83525  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-347 <STO>  
A:Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04360.1; GSPDB:GN001;  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: tolA; PA0971

Query Match 40.4%; Score 86; DB 2; Length 347;  
Best Local Similarity 51.1%; Pred. No. 0.33;  
Matches 24; Conservative 6; Mismatches 15; Indels 2; Gaps 1;

Qy 1 AKKYAKKAAEKAKKA--YKAAEAKKAAKYEKAAAEKAAAEKAAAEKAAAE 45  
Db 125 AQAQAAEKKADEAKKAAAEKAAAEQKKQADIAKKAEDAEAKKAAEDA 171

RESULT 15  
S61926  
histone H1 homolog - Bordetella pertussis  
C:Species: Bordetella pertussis  
C>Date: 23-Jul-1996 #sequence\_revision 06-Sep-1996 #text\_change 24-Nov-1999  
C:Accession: S61926; S69327  
R:Scarlato, V.; Arico, B.; Goyard, S.; Ricci, S.; Manetti, R.; Prugnola, A.; Manetti, R.  
Mol. Microbiol. 15, 871-881, 1995  
A:Title: A novel chromatin-forming histone H1 homologue is encoded by a dispensable and  
A:Reference number: S61926; MUID:95319329; PMID:7596289  
A:Accession: S61926  
A:Molecule type: DNA  
A:Residues: 1-182 <SCA1>  
A:Cross-references: EMBL:L37438; NID:g77717; PIDN:AAB59120.1; PID:g777718  
A:Accession: S69327  
A:Molecule type: protein  
A:Residues: 'X', 3-39 <SCA2>  
C:Genetics:  
A:Gene: bph1  
C:Superfamily: histone H1

Query Match 39.9%; Score 85; DB 2; Length 182;  
Best Local Similarity 61.4%; Pred. No. 0.25;  
Matches 27; Conservative 4; Mismatches 9; Indels 4; Gaps 3;

Qy 1 AKKYAKKAAEK--AKKAY-KAAEAKKAAKYEKAAAEKAAAEKAA 41  
Db 58 AKKVAKKAVAKKAVAKKAVAKKAVAKKAVA-KKAVAKKAVAKKA 100

Search completed: March 10, 2003, 12:28:26  
Job time : 14.1962 secs









```

RESULT 8
HBHA MYCTU
ID HBHA MYCTU STANDARD; PRT; 198 AA.
AC Q11142; O85733;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heparin-binding hemagglutinin (Adhesin).
GN HBHA OR RV0475 OR MT0493 OR MTCY20G9.01.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=H37RV, ATCC 201 / H37Ra, and BCG / Paris 1173 P2;
RX MEDLINE=98445421; PubMed=9770536;
RA Menozzi F.D., Bischoff R., Fort E., Brennan M.J., Loch C.;
RT "Molecular characterization of the mycobacterial heparin-binding
RT hemagglutinin, a mycobacterial adhesin.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12625-12630(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Barry C.E. III, Tekala F.,
RA Gordon S.V., Eiglmeier K., Gae S., Gars D., Chillingworth T., Connor R.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-16, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=ATCC 201 / H37Ra, and BCG / Paris 1173 P2;
RX MEDLINE=97188915; PubMed=9064359;
RA Menozzi F.D., Rouse J.H., Alavi M., Laude-Sharp M., Muller J.,
RA Bischoff R., Brennan M.J., Loch C.;
RT "Identification of a heparin-binding hemagglutinin present in
RT Mycobacteria.";
RL J. Exp. Med. 184:993-1001(1996).
RN [5]
RP FUNCTION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=103, and BCG;
RX MEDLINE=21342355; PubMed=11449276;
RA Pethé K., Alonso S., Biet F., Delogu G., Brennan M.J., Loch C.,
RA Menozzi F.D.;
RT "The heparin-binding haemagglutinin of M. tuberculosis is required for
RT extrapulmonary dissemination.";
RL Nature 412:190-194(2001).
CC -1- FUNCTION: REQUIRED FOR EXTRAPULMONARY DISSEMINATION. MEDIATES
CC ADHERENCE TO EPITHELIAL CELLS BY BINDING TO SULFATED
CC GLYCOCONGUATES PRESENT AT THE SURFACE OF THESE CELLS; BINDS
CC HEPARIN, DEXTRAN SULFATE, FUCOIDAN AND CHONDROITIN SULFATE.

```

```

CC PROMOTES HEMAGGLUTINATION OF ERYTHROCYTES OF CERTAIN HOST SPECIES.
CC INDUCES MYCOBACTERIAL AGGREGATION.
CC -1- SUBCELLULAR LOCATION: SURFACE ASSOCIATED.
CC -1- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL
CC DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL END
CC DIMINISH THE AFFINITY FOR HEPARIN.
CC -1- PTM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM
CC PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT
CC SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-
CC TERMINAL DOMAIN OF HBHA.
CC -1- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE
CC TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT
CC RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT
CC CONTAIN ANY.
CC -1- SIMILARITY: STRONG, TO M.LEPRAE HBHA.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF074390; AAC26052.1; -.
CC EMBL; Z77162; CAB00936.1; -.
CC TIGR; MT0493; -.
CC TubercuList; RV0475; -.
CC Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;
CC Virulence; Complete proteome.
CC INIT MET 0
CC DOMAIN 151 193 ALA/LYS-RICH.
CC CONFLICT 120 120 R -> P (IN REF. 1).
CC SEQUENCE 198 AA; 21403 MW; 513760F6F1EB6042 CRC64;
Query Match 40.4%; Score 86; DB 1; Length 198;
Best Local Similarity 65.6%; Pred. No. 0.068;
Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 11 EKAKYKAAEAKKAAKYEKAAEAKKAAEAKA 42
DB 160 KKAAPAKKAAKAAKAAKAAKAAKAAKAA 191
RESULT 9
MST1_DROHY
ID MST1_DROHY STANDARD; PRT; 344 AA.
AC Q08695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mst101(1).
GN MST101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=94200512; PubMed=8150205;
RA Neesen J., Buenemann H., Heinlein U.A.;
RT "The Drosophila hydei gene hmst101(1) encodes a testis-specific,
RT repetitive, axoneme-associated protein with differential abundance in
RT Y chromosome deletion mutant flies.";
RL Dev. Biol. 162:414-425(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC ASSOCIATED WITH AXONEMAL STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC SPERMATID BUNDLES.

```



CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
 CC -!- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X73480; CAA51875.1; -.  
 DR PIR; S34153; S34153.  
 DR FlyBase; FBgn0011816; Dhvdm101(1).  
 KW SpERM; Repeat; Multigene family.  
 FT DOMAIN 58 337 19 X 16 AA APPROXIMATE TANDEM REPEATS OF  
 FT K-K-K-X-X-E-X-A-[KQ]-K-X-X-E-X-A-X.  
 SQ SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;  
 Query Match 40.4%; Score 86; DB 1; Length 344;  
 Best Local Similarity 57.8%; Pred. No. 0.11;  
 Matches 26; Conservative 3; Mismatches 12; Indels 4; Gaps 2;  
 OY 2 KKYAKKAK-AEKKAKYKAAEAKK---AAKYKAAAEKAAAEKAA 42  
 DB 117 KEAAEKKKCAEAAKEAAEKKKCAEAAKEAAEKKKCAEAA 161  
 RESULT 10  
 TOLA\_PSEAE STANDARD; PRT; 347 AA.  
 AC P50600;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ToIA protein.  
 GN TOLA OR PA0971.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PAO;  
 RA Dennis J.J., Lafontaine E.R., Sokol P.A.;  
 RX MEDLINE=97113525; PubMed=8955385;  
 RT "Identification and characterization of the tolQRA genes of  
 RT Pseudomonas aeruginosa."  
 RL J. Bacteriol. 178:7059-7068(1996).  
 RN [2]  
 RP REVISIONS TO N-TERMINUS.  
 RA Duan K., Sokol P.A.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAO1;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 RT opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane  
 CC (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U39558; AAC44660.2; -.  
 DR EMBL; AE004530; AAG04360.1; -.  
 KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;  
 KW Complete proteome.  
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 17 37 POTENTIAL.  
 FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 209 216 POLY-ALA.  
 SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;  
 Query Match 40.4%; Score 86; DB 1; Length 347;  
 Best Local Similarity 51.1%; Pred. No. 0.11;  
 Matches 24; Conservative 6; Mismatches 15; Indels 2; Gaps 1;  
 OY 1 AKKYAKKAKAEKAAEAKK---YKAAEAKKAAKYKAAAEKAAAEKAAAE 45  
 DB 125 AOKAAEAKKAEKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAE 171  
 RESULT 11  
 DBH\_MYCBO STANDARD; PRT; 205 AA.  
 AC Q9XB18; Q9S5J5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).  
 GN HUP OR HLP OR MDP1.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AN5;  
 RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
 RX "HLPMT-A target for differentiation of M.tuberculosis and M.bovis."  
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BCG / Tokyo;  
 RA Matsumoto S., Yukitake H., Matsuo T., Minoda T., Yamada T.;  
 RX "Identification of a novel protein generating bacterial slow growth  
 RX from Mycobacterium."  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
 CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
 CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
 CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y18421; CAB46493.1; -.  
 DR EMBL; AB013441; BAA78330.1; -.  
 DR HSP; P02346; 1HUU.  
 DR InterPro; IPR000119; Bac\_DNABind.  
 DR InterPro; IPR001386; Histone\_H1/H5.  
 DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
 DR PRINTS; PR00624; HISTONEHS.  
 DR ProDom; PD000945; Bac\_DNABind; 1.



```

DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
KW DNA-binding; DNA condensation; Repeat.
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
FT CONFLICT 199 199 A -> T (IN REF. 2).
SQ SEQUENCE 205 AA; 21262 MW; 19FC67885DFE6A8 CRC64;

Query Match 40.1%; Score 85.5; DB 1; Length 205;
Best Local Similarity 58.5%; Pred. No. 0.078;
Matches 24; Conservative 2; Mismatches 12; Indels 3; Gaps 1;

Qy 1 AKYAKAKAKAKAYKAAEAKKAAKYEKAAEAKAAKEA 41
Db 111 AKKVAKKAPAKKATKAAKAAATKAPA---KKAATKAPAKKA 148

RESULT 12
H1_PARAN STANDARD; PRT; 248 AA.
AC P02256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal.
OS Parachinus angulosus (Angulate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinidae;
OC Parachinus.
OX NCBI_TaxID=7658;
RN [1]
RP SEQUENCE OF 1-84.
RX MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N.; Strickland M.; de Groot P.C.; von Holt C.;
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides.";
RL Eur. J. Biochem. 104:559-566(1980).
RN [2]
RP SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N.; Strickland M.; Brandt W.F.; von Holt C.; Lehmann A.;
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure.";
RL Eur. J. Biochem. 104:567-578(1980).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR; A02586; HSURIP.
DR HSSP; P02259; 1HST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; LinkerHist_N.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
FT VARIANT 144 144 K -> R.
SQ SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match 38.5%; Score 82; DB 1; Length 248;
Best Local Similarity 57.8%; Pred. No. 0.2;
Matches 26; Conservative 3; Mismatches 14; Indels 2; Gaps 2;

Qy 2 KKYAKKAAEAKKAYKAAEAKKAAKYEKAAEAKAAEAKAA 45
Db 120 KKAATKTAAPAKKA-KAAAKAKAKAKAAAKRRAALAKKAAAA 163

RESULT 13
H1E_CHIPA STANDARD; PRT; 235 AA.
AC P40262;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1E.
OS Chironomus pallidivittatus (Widge).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7151;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulze E.; Wisniewski J.R.; Nagel S.; Gavenis K.; Grossbach U.;
RA Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L29106; AAA21714.1; -.
CC HSSP; P08287; 1GHC.
CC InterPro; IPR001386; Histone H1/H5.
CC InterPro; IPR003216; LinkerHist_N.
CC Pfam; PF00538; linker histone; 1.
CC ProDom; PD000373; LinkerHist_N; 1.
CC SMART; SM00526; H15; 1.
CC Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 235 AA; 24446 MW; 340BC5B9A85002AC CRC64;

Query Match 38.0%; Score 81; DB 1; Length 235;
Best Local Similarity 41.3%; Pred. No. 0.23;
Matches 26; Conservative 2; Mismatches 17; Indels 18; Gaps 1;

Qy 1 AKYAKAKAAEAKA-----KKAYKAAEAKKAAKYEKAAEAKAAEAA 42
Db 126 AKKYDKPKKAPAPKPKSTNKRVTGKVVKKPAKKATKAAKPAKKVAAKPA 185

Qy 43 YEA 45
Db 186 KKA 188

RESULT 14
TOP1_STRCO STANDARD; PRT; 952 AA.
AC Q9X909;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA topoisomerase I (BC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Untwisting enzyme) (Swivelase).
GN TOPA OR SCO3543 OR SCH5.06C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;

```

```
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -!- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AL035636; CAB38480.1; -.
DR HSSP; P06612; 18CL.
DR InterPro; IPR002936; DNAPrim_toprim.
DR InterPro; IPR003601; DNATopI_ATP_bind.
DR InterPro; IPR003602; DNATopI_DNA_bind.
DR InterPro; IPR000380; Prok_topisomrase.
DR Pfam; PF01131; Topoisom_bac; 1.
DR Pfam; PF01751; Toprim_I.
DR PRINTS; PR00417; PRTPISMRASEI.
DR SMART; SM00437; TOPIAC; 1.
DR SMART; SM00436; TOP1BC; 1.
DR SMART; SM00493; TOPRIN; 1.
DR TIGRFAMs; TIGR01051; topA_bact; 1.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW Isomerase; Topoisomerase; DNA-binding; Complete proteome.
FT ACT_SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 952 AA; 103584 MW; 1C98DABD7A629295 CRC64;

Query Match 37.8%; Score 80.5; DB 1; Length 952;
Best Local Similarity 55.6%; Pred. No. 0.86; Mismatches 3; Gaps 2;
Matches 25; Conservative 2; Indels 15;

Oy 1 AKKYAKKA-KAEKAKK--AYKAAEAKKAAKYERKAAAEKAAAEKAA 42
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 888 AKTKAKVKKTAAKKAPAKKAAATKTKTAATKTAATKTAATKTA 932

RESULT 15
H1_SALT/
ID H1_SALT/ STANDARD; PRT; 194 AA.
AC P03254;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.
OS Salmo trutta (Brown trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_FaxID=8032;
RN [1]
```

```
RP SEQUENCE.
RX MEDLINE=78023898; PubMed=913397;
RA McLeod A.R., Wong N.C.W., Dixon G.H.;
RT "The amino-acid sequence of trout-testis histone H1.";
RL Eur. J. Biochem. 78:281-291(1977).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR; A02593; HSTR1.
DR HSSP; P08287; IGHC.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation; Phosphorylation.
FT MOD_RES 1 1 ACETYLATION (IN 90% OF THE CHAINS).
FT MOD_RES 145 145 PHOSPHORYLATION.
FT MOD_RES 161 161 PHOSPHORYLATION.
FT MOD_RES 182 182 PHOSPHORYLATION.
FT DOMAIN 27 100 GLOBULAR.
FT VARIANT 35 35 A -> I (IN MINOR COMPONENT).
SQ SEQUENCE 194 AA; 19408 MW; D288F9F44AF9BE7E CRC64;

Query Match 37.6%; Score 80; DB 1; Length 194;
Best Local Similarity 55.6%; Pred. No. 0.25;
Matches 25; Conservative 3; Mismatches 13; Indels 4; Gaps 2;

Oy 1 AKKYAKKA-KAEKAKK--AYKAAEAKKAAKY--EKAAAEKAAAEKAA 41
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 AKKPAKAAAPKAKKPAKAAKPKKPAKAAKPAKAAKPAKAAKPKKA 149

Search completed: March 10, 2003, 12:17:08
Job time : 11.2532 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 23.8291 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989a-2

Perfect score: 213

Sequence: 1 AKYAKKAKAKAKAYKAA.....AKYKAAAKAKAAKEAYEA 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	98	46.0	394	16 Q8X965	Q8X965 escherichia
2	94.5	44.4	239	16 Q8Y5W4	Q8Y5W4 listeria mo
3	91	42.7	244	16 Q9AJX2	Q9AJX2 streptomyce
4	88.5	41.5	243	16 Q92A67	Q92A67 listeria in
5	88.5	41.5	275	5 Q01395	Q01395 drosophila
6	88.5	41.5	372	2 Q9WMX1	Q9WMX1 pseudomonas
7	88	41.3	899	16 Q9A5J6	Q9A5J6 caulobacter
8	86.5	40.6	200	16 Q8XVN7	Q8XVN7 raietonia s
9	86	40.4	212	3 Q93946	Q93946 candida alb
10	85	39.9	98	5 Q8WQ44	Q8WQ44 leishmania
11	85	39.9	101	5 Q9BMV8	Q9BMV8 leishmania
12	85	39.9	111	5 Q8T9R3	Q8T9R3 leishmania
13	85	39.9	182	2 Q45370	Q45370 bordetella
14	84	39.4	60	5 Q9U3W3	Q9U3W3 caenorhabdi
15	83.5	39.2	483	12 Q8QND1	Q8QND1 ectocarpus
16	83	39.0	293	10 Q9AT18	Q9AT18 lens culina

17	83	39.0	389	16 Q9CM70	Q9CM70 pasteurella
18	82.5	38.7	243	5 Q23784	Q23784 chironomus
19	82	38.5	262	16 Q9S2M2	Q9S2M2 streptomyce
20	82	38.5	568	3 Q94567	Q94567 schizosacch
21	81.5	38.3	311	12 Q84528	Q84528 paramecium
22	81.5	38.3	482	2 Q93LK4	Q93LK4 enterococcu
23	81	38.0	81	5 Q9NFP6	Q9NFP6 trypanosoma
24	81	38.0	112	5 Q9XYV5	Q9XYV5 leishmania
25	81	38.0	241	5 Q23790	Q23790 chironomus
26	81	38.0	290	10 Q9AT24	Q9AT24 pisum sativ
27	81	38.0	295	10 Q9ZB20	Q9ZB20 pisum sativ
28	81	38.0	295	10 Q9AT22	Q9AT22 lathyrus sa
29	81	38.0	296	10 Q9AT25	Q9AT25 pisum sativ
30	81	38.0	297	10 Q9SXQ8	Q9SXQ8 pisum sativ
31	81	38.0	301	10 Q9AT23	Q9AT23 pisum sativ
32	81	38.0	306	10 Q9AT21	Q9AT21 lathyrus sa
33	80.5	37.8	277	16 Q9XAQ3	Q9XAQ3 streptomyce
34	80.5	37.8	445	3 Q9P3Q8	Q9P3Q8 neurospora
35	80	37.6	66	5 Q9S0Z0	Q9S0Z0 caenorhabdi
36	80	37.6	71	5 Q9NFX8	Q9NFX8 trypanosoma
37	80	37.6	81	5 Q9N6L9	Q9N6L9 trypanosoma
38	80	37.6	281	10 Q9AT20	Q9AT20 lens culina
39	80	37.6	291	16 Q9ZEM5	Q9ZEM5 listeria in
40	80	37.6	293	10 Q9AT19	Q9AT19 lens culina
41	80	37.6	347	16 Q9RKL9	Q9RKL9 streptomyce
42	80	37.6	969	5 Q9NDI9	Q9NDI9 plasmodium
43	80	37.6	1340	16 Q9L1H8	Q9L1H8 streptomyce
44	79.5	37.3	407	16 Q8ZQT6	Q8ZQT6 salmonella
45	79	37.1	62	5 Q17536	Q17536 caenorhabdi

#### ALIGNMENTS

#### RESULT 1

Q8X965 ID Q8X965 PRELIMINARY; PRT; 394 AA.  
AC Q8X965;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Membrane spanning protein, required for outer membrane integrity  
DE (Membrane spanning protein Tola)  
GN TOLA OR Z0907 OR ECS0774.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:111-22(2001).  
DR EMBL; AF005252; AAG55075.1; -.  
DR EMBL; AF002553; BAB34197.1; -.

DR InterPro; IPR00104; Antifreeze 1.  
 DR PRINTS; PR00308; ANTIFREEZE1.  
 KW Complete proteome.  
 SQ SEQUENCE 394 AA; 40517 MW; 5B58D8E230BDE28 CRC64;  
 Query Match 46.0%; Score 98; DB 16; Length 394;  
 Best Local Similarity 65.3%; Pred. No. 0.04; 9; Indels 6; Gaps 3;  
 Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;  
 Qy 2 KKYAKAKAEKAKK---KK--AYKAAEAKKAAKYEKAAAEKAAAEKAAAE 45  
 Db 220 KKAEEKAAAEKAAADKKAEEKAAADKKAEEKAAAEKAAAEKAAAE 267  
 ID Q8Y5W4 PRELIMINARY; PRT; 239 AA.  
 AC Q8Y5W4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein lmo1941.  
 GN LMO1941.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BGD-E / SEROVAR 1/2A;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Knecht J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 "Comparative genomics of Listeria species.";  
 Science 294:849-852(2001).  
 RL EMBL; AL591981; CAD00019.1; -;  
 DR ListiList; LMO01941; -;  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 DR SMART; SM00257; LysM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 239 AA; 25836 MW; 72E59D576E0D7832 CRC64;  
 Query Match 44.4%; Score 94.5; DB 16; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.053;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 Qy 2 KKYAKAKAEKAKK---AYKAAEAKK--AAKYEKAAAEKAAAEKAAAE 44  
 Db 124 KAAAEKAAEDKKQBEDAVKAAAEKAAAEKAAAEKAAAEKAAAE 171  
 ID Q9AJX2 PRELIMINARY; PRT; 244 AA.  
 AC Q9AJX2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative secreted protein.  
 GN SCO1905 OR SC133.04.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomyces.  
 OX NCBI\_TaxID=1639;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BGD-E / SEROVAR 1/2A;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Knecht J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 "Comparative genomics of Listeria species.";  
 Science 294:849-852(2001).  
 RL EMBL; AL591981; CAD00019.1; -;  
 DR ListiList; LMO01941; -;  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 DR SMART; SM00257; LysM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 239 AA; 25836 MW; 72E59D576E0D7832 CRC64;  
 Query Match 44.4%; Score 94.5; DB 16; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.053;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 Qy 2 KKYAKAKAEKAKK---AYKAAEAKK--AAKYEKAAAEKAAAEKAAAE 44  
 Db 124 KAAAEKAAEDKKQBEDAVKAAAEKAAAEKAAAEKAAAEKAAAE 171  
 ID Q9AJX2 PRELIMINARY; PRT; 244 AA.  
 AC Q9AJX2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative secreted protein.  
 GN SCO1905 OR SC133.04.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomyces.  
 OX NCBI\_TaxID=1639;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BGD-E / SEROVAR 1/2A;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Knecht J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 "Comparative genomics of Listeria species.";  
 Science 294:849-852(2001).  
 RL EMBL; AL591981; CAD00019.1; -;  
 DR ListiList; LMO01941; -;  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 DR SMART; SM00257; LysM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 239 AA; 25836 MW; 72E59D576E0D7832 CRC64;  
 Query Match 44.4%; Score 94.5; DB 16; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.053;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 Qy 2 KKYAKAKAEKAKK---AYKAAEAKK--AAKYEKAAAEKAAAEKAAAE 44  
 Db 124 KAAAEKAAEDKKQBEDAVKAAAEKAAAEKAAAEKAAAEKAAAE 171  
 ID Q9AJX2 PRELIMINARY; PRT; 244 AA.  
 AC Q9AJX2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative secreted protein.  
 GN SCO1905 OR SC133.04.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomyces.  
 OX NCBI\_TaxID=1639;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BGD-E / SEROVAR 1/2A;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Knecht J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 "Comparative genomics of Listeria species.";  
 Science 294:849-852(2001).  
 RL EMBL; AL591981; CAD00019.1; -;  
 DR ListiList; LMO01941; -;  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 DR SMART; SM00257; LysM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 239 AA; 25836 MW; 72E59D576E0D7832 CRC64;  
 Query Match 44.4%; Score 94.5; DB 16; Length 239;  
 Best Local Similarity 56.2%; Pred. No.



1

RESULT 14	
ID Q9U3W3	PRT; 60 AA.
AC Q9U3W3	PRELIMINARY;
DT 01-MAY-2000	(TREMBLrel. 13, Created)
DT 01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DE 01-MAR-2002	(TREMBLrel. 20, Last annotation update)
DE Histone H1-Q	(Hypothetical 6.2 kDa protein).
GN C01B10.5	
OS Caenorhabditis elegans.	
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;	
OC Rhabditiidae; Peloderinae; Caenorhabditis.	
OX NCBI_TaxID=6239;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Jedrusik M.A., Schulze E.;	
RT "The histone H1 complement of Caenorhabditis elegans.";	
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.	
[2]	
RP SEQUENCE FROM N.A.	
RC STRAIN=BRISTOL N2;	
EC MEDLINE=99069613; PubMed=9851916;	
RA None;	
RP "Genome sequence of the nematode C. elegans: a platform for	
RT investigating biology. The C. elegans Sequencing Consortium."	
RL Science 282:2012-2018(1998).	
[3]	
RP SEQUENCE FROM N.A.	
RC STRAIN=BRISTOL N2;	
RA Blanchard M., Bradshaw H.;	
RT "The sequence of C. elegans cosmid C01B10.";	
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.	
[4]	
RP SEQUENCE FROM N.A.	
RC STRAIN=BRISTOL N2;	
RA Waterston R.;	
RT "Direct Submission.";	
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.	
DR ENBL; AF216291; AAF23175.1; "-.	
DR ENBL; U58757; AAK66021.1; "-.	
DR InterPro; IPR001386; Histone_H1/H5.	
DR PRINTS; PR00624; HISTONEH5.	
KW Hypothetical protein.	
SQ SEQUENCE 60 AA; 6212 MW; D01ABB4CEC355566D CRC64;	
Query Match 39.4%; Score 84; DB 5; Length 60;	
Best Local Similarity 58.1%; Pred.No. 0.14;	
Matches 25; Conservative 4; Mismatches 12; Indels 2; Gaps	
Qy 2 KYAKKAAEKAKKAY-KAAEAKKA-AK-YEKAAAEKAAKEEA 42	
:	
Dd 14 KVAAPKPAPKPKKASPKKAAAPKAPKVPVKAAAKSPAKKAA 56	
RESULT 15	
Q8QND1	PRELIMINARY; PRT; 483 AA.
ID Q8QND1	
AC Q8QND1;	
DT 01-JUN-2002	(TREMBLrel. 21, Created)
DT 01-JUN-2002	(TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002	(TREMBLrel. 21, Last annotation update)
DE EsV-1-151.	
OS Ectocarpus siliculosus virus.	
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.	
OX NCBI_TaxID=37665;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ESV-1;	
RA Delaroque N.; Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W	
RT "The complete nucleotide sequence of the Ectocarpus siliculosus viru	
RT genome.";	





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 23.4494 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989a-2

Perfect score: 213

Sequence: 1 AKYAKAKAKAKAYKAA.....AKYKAAAKKAAKAAAYEA 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq 101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	45	21	AAV82572
2	138	64.8	109	21	AAV82577
3	134.5	63.1	56	21	AAV82573
4	131	61.5	77	21	AAV82575
5	126.5	59.4	86	21	AAV82576
6	120.5	56.6	66	21	AAV82574
7	108	50.7	35	21	AAV82571
8	99.5	46.7	106	11	AAV82571
9	98	46.0	46	13	AAV82871
10	94.5	44.4	239	23	ABR49123

11	94	44.1	154	11	AAV82572	Recombinant copoly
12	91	42.7	223	20	AAV14928	Amino acid sequenc
13	89.5	42.0	214	20	AAV34055	M. tuberculosis hi
14	89.5	42.0	214	21	AAV57353	M. tuberculosis hi
15	87	40.8	334	22	ABG28693	Novel human diagno
16	86	40.4	339	19	AAW44934	Mycobacterial hepa
17	86	40.4	198	19	AAW43082	Mycobacterial hepa
18	86	40.4	198	19	AAW44936	Mycobacterial hepa
19	85.5	40.1	205	21	AAV20575	Mycobacterium bovi
20	85	39.9	427	22	AAV70868	C albicans apoptos
21	82	38.5	165	22	AAV91997	C glutamicum prote
22	77.5	36.4	239	22	ABG63276	Drosophila melanog
23	77.5	36.4	2151	22	ABG60086	Drosophila melanog
24	77	36.2	103	20	AAV34067	Histone H1 isoform
25	77	36.2	103	21	AAV57365	Human histone H1.5
26	77	36.2	116	20	AAV34066	Histone H1 isoform
27	77	36.2	116	21	AAV57364	Human histone H1.5
28	77	36.2	158	20	AAV34068	Histone H1 isoform
29	77	36.2	158	21	AAV57366	Human histone H1.5
30	77	36.2	222	18	AAW29476	Human histone H1
31	77	36.2	222	20	AAV34033	Human histone H1
32	77	36.2	222	21	AAV57331	Human histone H1
33	77	36.2	226	20	AAV34060	Human histone H1
34	77	36.2	226	21	AAV57358	Human histone H1
35	76	35.7	157	22	ABG58855	Drosophila melanog
36	75.5	35.4	111	23	ABP00299	Human ORFX protein
37	75.5	35.4	201	23	ABP41465	Human ovarian anti
38	75.5	35.4	218	21	AAV00755	Human secreted pro
39	75.5	35.4	218	21	AAV00759	Human secreted pro
40	75.5	35.4	220	23	AAV76972	Human ribosomal L1
41	75.5	35.4	225	21	AAV00758	Human secreted pro
42	75.5	35.4	265	21	AAV58221	Lung cancer associ
43	75	35.2	160	18	AAV14549	Streptococcus pneu
44	74.5	35.0	32	16	AAV90180	Polyclonatic poly
45	74.5	35.0	32	18	AAV06686	Protamine-like pep

#### ALIGNMENTS

##### RESULT 1

AAV82572

ID AAV82572 standard; peptide; 45 AA.

XX AAV82572;

XX AC

XX DT 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:2.

XX Copolymer; molecular weight marker; TV-marker; immune disease;

XX glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;

XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;

XX antidiabetic; chymotrypsin; haemostatic; antipsoriatic; dermatological;

XX antianemic; immunosuppressive; demyelinating disease; osteoarthritis;

XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;

XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;

XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;

XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;

XX pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX OS

XX WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PA (TEVA-) TEVA PHARM USA INC.

XX	Gad A, Lis D;
PI	WPI; 2000-317499/27.
XX	Copolymer 1 related polypeptides used as molecular weight markers for
PT	glatiramer acetate and for treatment and prevention of immune diseases
XX	Claim 10; Page 14: 72pp; English.
XX	AAV82571 to AAV82577 represent specifically claimed copolymer molecular
CC	weight TV-marker polypeptides from the present invention. The present
CC	invention describes polypeptides (I) for determining the molecular
CC	weight of a copolymer (CP), which has an identified molecular weight
CC	and an amino acid composition corresponding to the copolymer. The
CC	polypeptides of the invention are used as molecular weight markers for
CC	glatiramer acetate related tetrapolymers. The polypeptides may also be
CC	used for treating and preventing immune diseases in a mammal. Autoimmune
CC	diseases which may be treated include either cell-mediated or
CC	antibody-mediated diseases. Such diseases include arthritic conditions,
CC	demyelinating diseases and inflammatory conditions, e.g. multiple
CC	sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC	anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC	uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC	purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC	disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC	myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC	lupus erythematosus. Mediated-mediated diseases which can be treated
CC	include host-versus-graft disease, graft-versus-host disease, and
CC	delayed-type hypersensitivity. The polypeptides of the invention have
CC	defined molecular weights and physical properties which are analogous to
CC	glatiramer acetate molecules, which makes them ideal for use as
CC	molecular weight markers.
XX	Sequence 45 AA;
SQ	
Query Match	100.0%; Score 213; DB 21; Length 45;
Best Local Similarity	100.0%; Pred. NO. 1.4e-16;
Matches 45;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 AKYAKAAGAEKAYKAABEAKKAATKEAAEKAAAEKAAEAEEA 45 
Dd	1 AKYAKAAGAEKAYKAABEAKKAATKEAAEKAAAEKAAEAEEA 45 
RESULT 2	
AAV82577	ID AAV82577 standard; peptide; 109 AA.
XX	AAV82577;
XX	28-JUL-2000 (first entry)
DE	Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:7.
XX	Copolymer; molecular weight marker; TV-marker; immune disease;
KW	glatiramer acetate; autoimmune disease; antiarthritis; neuroprotective;
KW	osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW	antiadrenergic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW	antianemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW	inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW	Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW	diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW	Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW	pemphigus vulgaris; systemic lupus erythematosus.
XX	Unidentified.
OS	
XX	WO200018794-A1.
PX	06-APR-2000.
PD	
FF	24-SEP-1999; 99WO-US22402.

XX	25-SEP-1998;	98US-0101693.
PR	(YEDA ) YEDA RES & DEV CO LTD.	
XX	(TEVA-) TEVA PHARM USA INC.	
XX		
XX	Gad A, Lis D;	
FI	WPI; 2000-317499/27.	
XX		
DR	Claim 10; Page 14; 72pp; English.	
XX		
XX	AAV82571 to AAV82577 represent specifically claimed copolymer molecular	
CC	weight TV-marker polypeptides from the present invention. The present	
CC	invention describes polypeptides (I) for determining the molecular	
CC	weight of a copolymer (CP), which has an identified molecular weight	
CC	and an amino acid composition corresponding to the copolymer. The	
CC	polypeptides of the invention are used as molecular weight markers for	
CC	glatiramer acetate related tetrapolymers. The polypeptides may also be	
CC	used for treating and preventing immune diseases in a mammal. Autoimmune	
CC	diseases which may be treated include either cell-mediated or	
CC	antibody-mediated diseases. Such diseases include arthritic conditions,	
CC	demylinating diseases and inflammatory conditions, e.g. multiple	
CC	sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic	
CC	anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune	
CC	uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia	
CC	purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves	
CC	disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic	
CC	myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic	
CC	lupus erythematosus. Mediated-mediated diseases which can be treated	
CC	include host-versus-graft disease, graft-versus-host disease, and	
CC	delayed-type hypersensitivity. The polypeptides of the invention have to	
CC	defined molecular weights and physical properties which are analogous to	
CC	glatiramer acetate molecules, which makes them ideal for use as	
CC	molecular weight markers.	
XX		
XX	Sequence 109 AA;	
XX		
XX	Query Match 64.8%; Score 138; DB 21; Length 109;	
XX	Best Local Similarity 76.6%; Pred. No. 5.1e-08;	
XX	Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps	
QY	1 AKKYAKKAAKAAKAYKAAEAKK--AAKYEKAAAEKAAKEAAVEA 45	
DB	65 AKKYAKKAAKAE--KKBYAAAEAKKAAKAAKAYKAAEAAKAAKEAAVEA 109	
XX		
XX	RESULT 3	
XX	AAV82573	
ID	AAV82573 standard; peptide; 56 AA.	
XX		
XX	AAV82573;	
XX		
DT	28-JUL-2000 (first entry)	
XX		
DE	Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.	
XX		
KW	Copolymer; molecular weight marker; TV-marker; immune disease;	
KW	glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;	
KW	osteopathic; immunosuppressive; antithyroid; antiinflammatory;	
KW	antidiabetic; thymostatic; haemostatic; antipsoriatic; dermatological;	
KW	antianemic; immunosuppressive; demyelinating disease; osteoarthritis;	
KW	inflammatory condition; multiple sclerosis; rheumatoid arthritis;	
KW	Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;	
KW	diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;	
KW	Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;	
KW	pemphigus vulgaris; systemic lupus erythematosus.	
XX		
XX	Unidentified.	
XX		







Qy	3	KYAKKA-KAEKAKAYKAAEAKKAAKYKAAAEKAA-----AKEA	41
		:         :   :	
Db	8	KAACKAYAEKAKAKYKAEKAKAEKAAEAKKAAEAKKAAEAKKAAEAKA	55
RESULT 9			
AAAR28871		AAAR28871 standard; peptide; 46 AA.	
XX	AAAR28871;		
XX	23-MAR-1993	(first entry)	
DE	High affinity macrophage mannose receptor ligand compound #9.		
XX	glycopeptide; mannose; mannosylated; glycosylated; mannose receptor;		
KW	macrophages; monocytes; destroy; cytotoxicity; label; image; alter;		
KW	macrophage processing of antigen; MHC restriction; inflammation;		
KW	inflammatory diseases; macrophage secretory products;		
KW	Crohn's disease; legionnaires disease; mononuclear phagocytes; HIV;		
KW	AIDS; lysosomal storage diseases; Gaucher's disease; asthma;		
KW	alveolar macrophages metastasis; systemic macrophages; deliver;		
KW	antigenic peptides; prevent transplant rejection;		
KW	organ transplantation; antitumour agents; cancer; toxins.		
XX	Synthetic.		
OS			
XX	Key	Location/Qualifiers	
PH	Modified-site 1		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine. May also have non interfering substituents."	
FT	Modified-site 4		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 7		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 10		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 13		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 16		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 19		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 22		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 25		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 28		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 31		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 34		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 37		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 40		
FT	/note=	"opt may have mannose, fucose, glucose or N-Ac-glucosamine."	
FT	Modified-site 43		

XX PD 18-OCT-2001.  
 XX XX  
 XX PF 11-APR-2001; 2001WO-FR011118.  
 XX PR 11-APR-2000; 2000FR-0004629.  
 XX PA (INSP ) INST PASTEUR.  
 XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Coseart P;  
 PI Daniels J, Coehel W, Krsft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX DR WPI; 2002-010914/01.  
 XX PS Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides -  
 XX Claim 6; SEQ ID No 1828; 192pp; French.  
 XX CC The present invention relates to the genome sequence of *Listeria*  
 CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC *monocytogenes* and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L. monocytogenes* and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by *L.*  
 CC *monocytogenes* and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 239 AA;  
 Query Match 44.4%; Score 94.5; DB 23; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.0061;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 QY 2 KKYAKKAAKAKK--AYKAAEAKK--AAKYKAAAEKAAKAAEAYE 44  
 Db 124 KAAAEKAEADKKKQFEEDAVKAAKAKKQEEAEKAAAEKAAAEKAAAE 171  
 RESULT 11  
 AAR06445  
 ID AAR06445 standard; protein; 154 AA.  
 XX AC  
 XX AC AAR06445;  
 XX DT 03-JAN-1991 (first entry)  
 XX DE Recombinant copolymer 1-77, myelin basic protein analogue.  
 XX XX Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
 KW immunological activity; autoimmune encephalomyelitis;  
 KW multiple sclerosis;  
 KW Synthetic.  
 OS  
 XX EF383620-A.  
 XX PN

XX PD 22-AUG-1990.  
 XX XX  
 XX PF 16-FEB-1990; 90EP-0301700.  
 XX PR 07-FEB-1990; 90US-0473845.  
 XX PR 17-FEB-1989; 89US-0312541.  
 XX PA (REPL-) REPLIGEN CORP.  
 XX PI Cook KS;  
 XX XX WPI; 1990-255848/34.  
 DR N-PSDB; AAQ05664.  
 XX PT Producing genes encoding random polymers of aminoacid(s) - for  
 PT producing recombinant polypeptide(s) with biological and/or  
 PT immunological activity  
 XX PS Disclosure; Fig 11; 25pp; English.  
 XX CC To improve the expression of rCOP-1 polypeptides in *E. coli*, genes  
 CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaTAN  
 CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
 CC express Protein A. The resulting plasmids encode fusion proteins  
 CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
 CC A methionine residue occurs between the Protein A and rCOP-1  
 CC sequences, originating from the 5' linker sequence, in order that  
 CC the COP-1 polypeptide may be cleaved from the fusion protein.  
 CC rCOP-1-77 contains oligonucleotide duplexes encoding the following  
 CC segments: YKK, EAE, KAK, AAA. The N-terminal alanine residue  
 CC is left behind following CNBr cleavage of the fusion protein.  
 CC The product prevents or arrests experimental autoimmune  
 CC encephalomyelitis. They are used to prevent, arrest or control a  
 CC demyelinating disorder, e.g. multiple sclerosis. They may also  
 CC be used as additives to hair care products to confer beneficial  
 CC effects on damaged hair or as supplements for diets deficient in  
 CC certain amino acids.  
 CC See also AAQ05665.  
 XX SQ Sequence 154 AA;  
 Query Match 44.1%; Score 94; DB 11; Length 154;  
 Best Local Similarity 60.5%; Pred. No. 0.0044;  
 Matches 26; Conservative 2; Mismatches 9; Indels 6; Gaps 2;  
 QY 2 KKYAKKAAKAKKAYKAAEAKKAAKYKAAAEKAAAEKAAEAYE 44  
 Db 105 KKYKKKAKKAKYKK--KAKEAKA-----KAAAEAKKAEAYK 141  
 RESULT 12  
 AAY14928  
 ID AAY14928 standard; protein; 223 AA.  
 XX AC  
 XX AC AAY14928;  
 XX DT 25-OCT-1999 (first entry)  
 XX DE Amino acid sequence of *M. vaccae* antigen GV-45.  
 XX KW Mycobacterium *vaccae* protein; antigen; T cell activation; cytokine;  
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;  
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;  
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;  
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;  
 KW squamous cell carcinoma; melanoma.  
 XX OS Mycobacterium *vaccae*.  
 XX XX WO9932634-A2.  
 XX XX 01-JUL-1999.  
 XX PD









THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	213	100.0	45	10	US-09-816-989A-2	Sequence 2, Appli
2	138	64.8	109	10	US-09-816-989A-7	Sequence 7, Appli
3	134.5	63.1	56	10	US-09-816-989A-3	Sequence 3, Appli
4	131	61.5	77	10	US-09-816-989A-5	Sequence 5, Appli
5	126.5	59.4	86	10	US-09-816-989A-6	Sequence 6, Appli
6	120.5	56.6	66	10	US-09-816-989A-4	Sequence 4, Appli
7	108	50.7	35	10	US-09-816-989A-1	Sequence 1, Appli
8	91	42.7	223	9	US-10-051-643-201	Sequence 201, App
9	82	38.5	165	9	US-09-738-626-5751	Sequence 5751, App
10	78	36.6	309	9	US-09-820-843A-24	Sequence 24, Appli
11	75.5	35.4	220	10	US-09-923-304-2	Sequence 2, Appli
12	75.5	35.4	265	10	US-09-925-302-559	Sequence 559, App
13	74	34.7	342	9	US-09-820-843A-23	Sequence 23, Appli
14	74	34.7	356	9	US-09-820-843A-27	Sequence 27, Appli
15	72.5	34.0	372	9	US-09-820-843A-8	Sequence 8, Appli
16	70.5	33.1	619	9	US-09-882-774-1	Sequence 1, Appli
17	70	32.9	641	10	US-09-765-272-160	Sequence 160, App
18	70	32.9	1156	10	US-09-815-242-13187	Sequence 13187, A
19	68	31.9	241	10	US-09-938-803-8	Sequence 8, Appli

; APPLICANT: Gad, Alexander  
: APPLICANT: Lis Davis

```
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1998-09-25
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match      64.8%; Score 138; DB 10; Length 109;
Best Local Similarity 76.6%; Pred. No. 3.6e-08;
Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps 2;

Qy 1 AKYAKAAEAKKAYKAAEAKK--AAKYKAAAEKAAAEKAAEAYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 AKYAKAAAE--KKEVAAAEAKKAAKAYKAAEAKKAAAEKAAEAYEA 109

RESULT 3
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

Query Match      63.1%; Score 134.5; DB 10; Length 56;
Best Local Similarity 68.4%; Pred. No. 4.2e-08;
Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAEAKK--AAKYKAAAEKAAAEKAAEAYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAYEA 56

RESULT 4
US-09-816-989A-5
; Sequence 5, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
```

```
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match      61.5%; Score 131; DB 10; Length 77;
Best Local Similarity 50.6%; Pred. No. 1.4e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE----- 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAYEA 60

Qy 30 -KAAAEKAAAEKAAEAYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 YKAEAAKAAAEKAAEAYEA 77

RESULT 5
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match      59.4%; Score 126.5; DB 10; Length 86;
Best Local Similarity 45.3%; Pred. No. 4.5e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 41; Gaps 3;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE----- 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAYEA 86

Qy 30 -----KAAAEKAAAEKAAEAYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 YKAEAAKAAAEKAAAEKAAAEKAAEAYEA 86

RESULT 6
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
```

```
; TITLE OF INVENTION:  of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION:  System using Mycobacterium Vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-201

Query Match          42.7%; Score 91; DB 9; Length 223;
Best Local Similarity 59.8%; Pred.No. 0.0059;
Matches 28; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY      1 AKKVA-KKAKEAKKAYKAAEKKAAKYEKAA-AEKAAAEEAAYEA 45
       ||| ||||| :||| ||||| :||| ||||| :||| ||||| :
Db     137 AKAATTAAPAKPAKATAAKKAPAAPKATAAKAAPAKKAPAKPAKAAATKA 183

RESULT 9
US-09-738-626-5751
; Sequence 5751, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAL, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5751
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5751

Query Match          38.5%; Score 82; DB 9; Length 165;
Best Local Similarity 62.9%; Pred.No. 0.038;
Matches 22; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY      7 KAKEAEKAKYKAAEKKAAKYEKAAEKKAAEKKAAEKA 41
       :||| :||| :||| :||| :||| :||| :||| :||| :
Db     120 EAITEKKKKAREDKAEKAA--EKAAAAEKAASAES 152

RESULT 10
US-09-820-843A-24
; Sequence 24, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: polyhydroxyalkanoate synthesis protein Phaf
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|9951352
US-09-820-843A-24

Query Match          36.6%; Score 78; DB 9; Length 309;
Best Local Similarity 54.8%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1 AKYAKKAAEKAKKAYKAAEAKEAAKAYEKAAAEKAAAEKAAAEKAA 42
Db 142 AKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 183

RESULT 11
US-09-923-304-2
; Sequence 2, Application US/09923304
; Patent No. US20020081612A1
; GENERAL INFORMATION:
; APPLICANT: KATZ, RUTH
; APPLICANT: JIANG, FENG
; TITLE OF INVENTION: DETECTION AND DIAGNOSIS OF SMOKING RELATED CANCERS
; FILE REFERENCE: UTSC:658US
; CURRENT APPLICATION NUMBER: US/09/923,304
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-304-2

Query Match          35.4%; Score 75.5; DB 10; Length 220;
Best Local Similarity 51.0%; Pred. No. 0.24;
Matches 25; Conservative 7; Mismatches 10; Indels 7; Gaps 3;

Qy 1 AKYAKKAAEKAKKAYKAAEAKEAAKAYEKAAAEKAAAEKAAAEKAA 42
Db 168 AKKTAASKKAPAKVPAQKATGQKAAPAKPAKQKQKAPAKAPKAPKAS 216

RESULT 12
US-09-925-302-559
; Sequence 559, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL04
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 559
```

```
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-559

Query Match          35.4%; Score 75.5; DB 10; Length 265;
Best Local Similarity 51.0%; Pred. No. 0.29;
Matches 25; Conservative 7; Mismatches 10; Indels 7; Gaps 3;

Qy 1 AKKY---AKKAAEK--AKKA--YKAAEAKEAAKAYEKAAAEKAAAEKAA 42
Db 213 AKKTAASKKAPAKVPAQKATGQKAAPAKPAKQKQKAPAKAPKAPKAS 261

RESULT 13
US-09-820-843A-23
; Sequence 23, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: algininate regulatory protein AlgP
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|9951563
US-09-820-843A-23

Query Match          34.7%; Score 74; DB 9; Length 352;
Best Local Similarity 51.1%; Pred. No. 0.56;
Matches 23; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 AKYAKKAAEKAKKAYKAAEAKEAAKAYEKAAAEKAAAEKAAAEKAA 45
Db 160 AKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 204

RESULT 14
US-09-820-843A-27
; Sequence 27, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: tola protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|9656364
```

Job time : 13.5289 secs

**THIS PAGE BLANK (USPTO)**





APPLICATION NUMBER: US/09/041,8

```

Query Match          36.2%; Score 77; DB 3; Length 103;
Best Local Similarity 52.2%; Pred. No. 0.068;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKAYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKKAVKKTTPKAKKPAA 96

RESULT 5
US-09-041-889-38
; Sequence 38, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 41
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-38

Query Match          36.2%; Score 77; DB 3; Length 116;
Best Local Similarity 52.2%; Pred. No. 0.077;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKAYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKKAVKKTTPKAKKPAA 96

RESULT 6
US-09-041-889-40
; Sequence 40, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using

```

```

; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 40
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-40

Query Match          36.2%; Score 77; DB 3; Length 158;
Best Local Similarity 52.2%; Pred. No. 0.11;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKAYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKKAVKKTTPKAKKPAA 96

RESULT 7
US-09-041-889-3
; Sequence 3, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..222
; OTHER INFORMATION: /note= "product = Human Histone
; OTHER INFORMATION: H1-S-3"
US-09-041-889-3

Query Match 36.2%; Score 77; DB 3; Length 222;
Best Local Similarity 52.2%; Pred. No. 0.15;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKAYKAAEAKKAAKYEKAAAEKAAA 38
Db 118 AKPKAKKAGAAKAKKAGATPKKAKKAGAKKAVKKTPTKKAKKPA 163

RESULT 9
US-09-041-889-32
; Sequence 32, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-041-889-32

Query Match 36.2%; Score 77; DB 3; Length 226;
Best Local Similarity 52.2%; Pred. No. 0.15;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKAYKAAEAKKAAKYEKAAAEKAAA 38
Db 119 AKPKAKKAGAAKAKKAGATPKKAKKAGAKKAVKKTPTKKAKKPA 164

RESULT 10
US-08-152-488-13
; Sequence 13, Application US/08152488
; Patent No. 5534619
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
```

```

Query Match      35.0%; Score 74.5; DB 1; Length 32;
Best Local Similarity 68.8%; Pred. No. 0.038;
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

QY 5 AKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKA 36
||||| : ||||| : ||||| : ||||| :
Db 2 AKKA-AKKAKKA--AKKAKKAAKKAKKAAKKA 30
||||| : ||||| : ||||| : ||||| :

RESULT 12
US-08-677-304-13
Sequence 13. Application US/08677304

```

RESULT 12  
US-08-677-304-13  
; Sequence 13, Application US/08677304  
; Patent No 5721212

; PATENT NO. 5/21212  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: Wakefield, Thomas W.  
 ;  
 ; APPLICANT: Andrews, Philip C.  
 ;  
 ; APPLICANT: Stanley, James C.

TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
 TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Benita J. Rohm, Esq.  
 STREET: 512 Springfield Avenue  
 CITY: Cranford  
 STATE: New Jersey  
 COUNTRY: United States of America  
 ZIP: 07016-1811  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Wordperfect 6; ASCII (DOS)Text

RESULT 13  
 US-08-436-703B-2  
 Sequence 2, Application US/08436703B  
 Patent No. 5919761  
 GENERAL INFORMATION:  
 APPLICANT: Wakefield, Thomas W.  
 APPLICANT: Andrews, Philip C.  
 APPLICANT: Stanley, James C.  
 TITLE OF INVENTION: NOVEL PEPTIDES FOR  
 TITILE OF INVENTION: HEPARIN AND LOW MOLECULAR  
 TITILE OF INVENTION: WEIGHT HEPARIN  
 TITILE OF INVENTION: ANTICOAGULATION REVERSAL  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Benita J. Rohm, Esq.  
 STREET: 6601 Woodward Avenue  
 STREET: Suite 1525  
 CITY: Detroit  
 STATE: Michigan  
 COUNTRY: United States of America  
 ZIP: 48226  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk 1.44MB, 3.5"  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WordPerfect 6:  
 SOFTWARE: ASCII (DOS)Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/436,703B  
 FILING DATE: 08-MAY-1995  
 CLASSIFICATION: 514

```

RESULT 14
US-08-303-025-16
; Sequence 16, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: Wordperfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-7622
; TELEFAX: 313-496-8454
; INFORMATION FOR SEQ ID NO. 16:

```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-16

Query Match 35.0%; Score 74.5; DB 1; Length 33;
Best Local Similarity 68.8%; Pred. No. 0.039;
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

Oy 5 AKKAKEAKKAYKAAEAKKAAKYEKAAAEKA 36
    ||||| :||||| :||||| :||||| :|||
Db 3 AKKA-AKKAKKA--AKKAKKAAKKAKKAAKKA 31

Search completed: March 10, 2003, 12:30:02
Job time : 8.97468 secs

; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; US-08-436-703B-4

Query Match 35.0%; Score 74.5; DB 2; Length 33;
Best Local Similarity 68.8%; Pred. No. 0.039;
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

Oy 5 AKKAKEAKKAYKAAEAKKAAKYEKAAAEKA 36
    ||||| :||||| :||||| :||||| :|||
Db 3 AKKA-AKKAKKA--AKKAKKAAKKAKKAAKKA 31

Search completed: March 10, 2003, 12:30:02
Job time : 8.97468 secs

; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-16

Query Match 35.0%; Score 74.5; DB 1; Length 33;
Best Local Similarity 68.8%; Pred. No. 0.039;
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

Oy 5 AKKAKEAKKAYKAAEAKKAAKYEKAAAEKA 36
    ||||| :||||| :||||| :||||| :|||
Db 3 AKKA-AKKAKKA--AKKAKKAAKKAKKAAKKA 31

RESULT 15
US-08-436-703B-4
; Sequence 4, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 6601 Woodward Avenue
; STREET: Suite 1525
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk 1.44Mb, 3.5"
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 6;
; SOFTWARE: ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,703B
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: N/A
; FILING DATE: N/A
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: 7WK-060548-00233
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-965-1976
; TELEFAX: 313-965-1951
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 16.4219 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKKAAK.....EAKYKAEAKAAKAAEAYEA 56  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	117	44.0	421	2 JV0057	tolA protein - Esc
2	116.5	43.8	239	2 AE1317	hypothetical prote
3	113	42.5	394	2 F90725	membrane spanning
4	113	42.5	394	2 G85576	membrane spanning
5	110.5	41.5	376	2 AG0592	tolA protein [impo
6	110	41.4	1701	2 T09127	probable erythrocy
7	105.5	39.7	372	2 G64064	outer membrane int
8	102.5	38.5	210	2 A25550	histone H1 - sea u
9	102.5	38.5	347	2 E83525	histone H1 - sea u
10	102	38.3	211	2 A28100	tolA protein PA097
11	100.5	37.8	1306	2 S51364	histone H1-beta, e
12	99.5	37.4	206	1 HSTR1R	sperm tail-specifi
13	98.5	37.0	243	2 AE1689	histone H1 - rainb
14	98.5	37.0	311	2 T17698	hypothetical prote
15	97.5	36.7	206	2 S09388	hypothetical prote
16	97	36.5	219	2 B60110	histone H1 - sea u
17	96	36.1	384	2 B43592	repetitive protein
18	95.5	35.9	217	2 A26721	outer membrane pro
19	95.5	35.9	248	1 HSUR1P	histone H1-gamma,
20	95.5	35.9	1403	2 T11583	probable translati
21	95	35.7	388	2 AC0138	tolA colicin impor
22	94.5	35.5	214	2 G70673	probable hupB - My
23	94.5	35.5	328	2 A44993	cytosolic repetiti
24	94	35.3	433	2 S25194	zucotin - yeast (Sa
25	94	35.3	703	2 T48600	kinase-like protei
26	92.5	34.8	182	2 S61926	histone H1 homolog
27	92.5	34.8	1128	2 T30296	R27-2 protein - Tr
28	92	34.6	220	2 A28456	histone H1.10 - ch
29	92	34.6	236	2 S22322	histone H1 - wheat

histone H1-delta -  
histone H1 (clone  
histone H1-II - Vo  
conserved hypochet  
Iga-specific metal  
histone H1 - Chlam  
immediate-early pr  
translation initia  
probable late embr  
histone H1 - trout  
tolA protein VC183  
arylesterase-relat  
hypothetical prote  
hypothetical prote  
histone H1 (clone  
hypothetical prote

ALIGNMENTS

RESULT 1

JV0057

tolA protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002

C;Accession: JV0057; B64810

R;Levensgood, S.K.; Webster, R.E.

J. Bacteriol. 171, 6600-6609, 1989

A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their product

A;Reference number: JV0057; MUID:90078104; PMID:2687247

A;Accession: JV0057

A;Molecule type: DNA

A;Residues: 1-421 <LEV>

A;Cross-references: GB:M28232; NID:gl48018; PIDN:AAA24683.1; PID:gl48019

A;Experimental source: strain JM105

A;Note: the authors translated the initiation codon GTG for residue 1 as Val

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64810

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <BLAT>

A;Cross-references: GB:AE000177; GB:U00096; NID:gl786955; PIDN:AACT3833.1; PID:gl786960;

A;Experimental source: strain K-12, substrain MG1655

C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach t

C;Genetics:

A;Gene: tolA

A;Map position: 17 min

A;Start codon: GTG

C;Keywords: nucleotide binding; P-loop; transmembrane protein

F;14-34/Domain: transmembrane #status predicted <MSS>

F;78-301/Domain: helical #status predicted <HSR>

F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 44.0%; Score 117; DB 2; Length 421;

Best Local Similarity 54.7%; Pred. No. 0.0043;

Matches 35; Conservative 5; Mismatches 10; Indels 14; Gaps 2;

QY 7 KEKAYAKKAAKAAKAAKAYK-----AAEAKKAAEA-----KYKAAKAAKAAKAA 52

Db 148 KAEADAKAAEAAKAAADAKKAAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAAEAA 207

QY 53 AYE 56

Db 208 RKKA 211

RESULT 2

AE1317



**Qy** 2 KKYAKKEKAYAKAEKAAKAEKAYYAAEAK-----KKAEAKYKA 42  
| | : | : ||| : |||| | | : ||| :  
**Dd** 96 KTRAQKARAAAKKLAAKKEQEKEKAAKTARKEKLAAKGAAKAAKKVKKPPAAKAK 155





THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 12.7595 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYKAEAKAAKAEAYEA 56

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	44.0	421	1	TOLA_ECOLI
2	105.5	39.7	372	1	TOLA_HAEN
3	105	39.5	349	1	RS6_AEDAL
4	104.5	39.3	346	1	RS6_AEDAE
5	102.5	38.5	210	1	H1_LYTP1
6	102.5	38.5	347	1	TOLA_PSEAE
7	100.5	37.8	1391	1	MST2_DROHY
8	99.5	37.4	206	1	H1_OHCNY
9	99.5	37.4	237	1	H1E_CHITE
10	99	37.2	211	1	H1B_STRPU
11	97.5	36.7	205	1	DBH_MYCBO
12	96	36.1	384	1	TMPE_TREPH
13	95.5	35.9	217	1	H1G_STRPU
14	95.5	35.9	248	1	H1_PARAN
15	95.5	35.9	1403	1	YDF3_SCHPO
16	95	35.7	917	1	IF2_PROVU
17	94.5	35.5	214	1	DBH_MYCTU
18	94	35.3	433	1	ZUO1_YEAST
19	93.5	35.2	208	1	DBH_MYCSM
20	92	34.6	219	1	H1O_CHICK
21	91.5	34.4	185	1	H1D_STRPU
22	91.5	34.4	238	1	H1_WHEAT
23	91	34.2	139	1	ASR_KLEPN
24	91	34.2	221	1	H1C_CHITE
25	91	34.2	240	1	H12_VOLCA
26	90.5	34.0	235	1	H1E_CHIPA
27	90.5	34.0	1532	1	IGA_NEIGO
28	90	33.8	407	1	IE68_HSVSA
29	90	33.8	884	1	IF2_YERPE
30	89.5	33.6	194	1	H1_SALTR
31	88.5	33.3	233	1	H1I_GLYSA
32	88	33.1	771	1	CALD_CHICK
33	87.5	32.9	233	1	H1I_GLYBA

## RESULT 1

ID	TOLA_ECOLI	STANDARD;	PRT;	421 AA.
AC	P19934;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	TOLA protein.			
GN	TOLA OR CIM OR EXCC OR LKY OR B0739.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JM105;			
RX	MEDLINE=90078104; PubMed=2687247;			
RA	Levengood S.K., Webster R.E.;			
RT	"Nucleotide sequences of the tola and tolB genes and localization of their products, components of a multistep translocation system in Escherichia coli."			
RT	J. Bacteriol. 171:6600-6609(1989).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;			
RA	"The complete genome sequence of Escherichia coli K-12."			
RT	Science 277:1453-1474(1997).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97061202; PubMed=8905232;			
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;			
RA	"A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map."			
RT	DNA Res. 3:137-155(1996).			
RL	[4]			
RN	DOMAINS.			
RP	MEDLINE=91296736; PubMed=2068069;			
RX	Levengood S.K., Beyer W.F. Jr., Webster R.E.;			
RA	"Tola: a membrane protein involved in colicin uptake contains an extended helical region."			
RT	Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).			
RL	[5]			
RN	INTERACTION WITH PORINS.			
RP	MEDLINE=97133271; PubMed=8978668;			
RX	Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,			

Q08695 drosophila  
P08284 gallus gall  
O86537 streptomyc  
P40276 chironomus  
P19649 treponema p  
Q02281 chlanydia t  
Q02257 echinolaipa  
P02987 gallus gall  
Q07134 chironomus  
Q9x909 streptomyc  
O9xq15 bacteroides  
P10771 caenorhabdi

## ALIGNMENTS

```

RA Llobes R.;
RT "TOLA central domain interacts with Escherichia coli porins.";
RL EMO J. 15:6408-6415 (1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX MEDLINE=99332679; PubMed=10404600;
RA Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;
RT "Filamentous phage infection: crystal structure of g3p in complex
with its coreceptor, the C-terminal domain of TOLA.";
RL Structure 7:711-722 (1999).
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION
OF BACTERIOPHAGE DNA.
CC -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
AND LAMB.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M28232; AAA24683.1; -.
DR EMBL; AE000177; AAC73833.1; -.
DR EMBL; D90713; BAA35405.1; -.
DR PIR; JVO057; JVO057.
DR PDB; 1TOL; 20-MAY-99.
DR EcoGene; EG11007; tola.
KW Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW Repeat; Inner membrane; 3D-structure; Complete proteome.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34
FT DOMAIN 35 421 PERIPLASMIC (POTENTIAL).
FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).
FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).
FT DOMAIN 224 278 10 X TANDEM REPEATS OF [ED]-K(1,2) -
A(2,4).
FT SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;
Query Match 44.0%; Score 117; DB 1; Length 421;
Best Local Similarity 54.7%; Pred. NO. 0.0024;
Matches 35; Conservative 5; Mismatches 10; Indels 14; Gaps 2;
Qy 7 KEKAYAKAEAKKAEAKAYK-----AAEAKKAE-----KYKAEAKAAKAE 52
Db 148 KAEADAKAEAEAKKAADAKKAEAEAEAKKAEAEAKKAEAEAEAKKAEAEAEAE 207
Qy 53 AYE 56
Db 208 RKKA 211
RESULT 2
TOLA HAETN STANDARD; PRT; 372 AA.
AC P44678; P94810;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE TOLA protein.
GN TOLA OR H10383.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RS6 / KW20 / ATCC 51907;

```

```

RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.W.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1479;
RX MEDLINE=97080550; PubMed=8921895;
RA Sen K., Sikkema D.J., Murphy T.F.;
RT "Isolation and characterization of the Haemophilus influenzae toIQ,
RT toLR, toLA and toLB genes.";
RL Gene 178:75-81 (1996).
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
(BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
(Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U32722; AAC22041.1; -.
DR EMBL; U32470; AAC44596.1; -.
DR HSSP; P19934; 1TOL.
DR TIGR; H10383; -.
KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 9 29
FT DOMAIN 30 372 PERIPLASMIC (POTENTIAL).
FT VARIANT 48 48 V -> A (IN STRAIN 1479).
FT VARIANT 142 142 K -> R (IN STRAIN 1479).
FT VARIANT 165 165 A -> P (IN STRAIN 1479).
FT VARIANT 190 190 A -> R (IN STRAIN 1479).
FT VARIANT 203 203 V -> A (IN STRAIN 1479).
FT VARIANT 227 227 D -> A (IN STRAIN 1479).
FT VARIANT 232 232 A -> AKAAAKAKA (IN STRAIN 1479).
FT VARIANT 234 234 T -> A (IN STRAIN 1479).
FT VARIANT 249 249 L -> F (IN STRAIN 1479).
FT VARIANT 254 254 I -> V (IN STRAIN 1479).
FT VARIANT 306 306 N -> S (IN STRAIN 1479).
FT VARIANT 323 323 T -> A (IN STRAIN 1479).
FT VARIANT 333 333 S -> P (IN STRAIN 1479).
FT SEQUENCE 372 AA; 39831 MW; 266ECF05C6C95544 CRC64;
Query Match 39.7%; Score 105.5; DB 1; Length 372;
Best Local Similarity 48.2%; Pred. NO. 0.019;
Matches 27; Conservative 14; Mismatches 14; Indels 1; Gaps 1;
Qy 2 KYAKKAEKAYAKAEKAAK-KAEAKYKAAAEAKKAEAKYKAEAKKAEAKKAEAKAE 56
Db 133 EKQQAEEAKAKQAEEAAKLKADAEAKRLAAAKQAEEAEKAEAKAEAKAE 189
RESULT 3
RS6_AEDAL
ID RS6_AEDAL STANDARD; PRT; 349 AA.
AC Q90762;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```









FT	DOMAIN	1	90	BACTERIAL HISTONE-LIKE DOMAIN.
FT	DOMAIN	101	205	DEGENERATE REPEATS REGION.
FT	CONFLICT	199	199	A -> T (IN REF. 2).
SQ	SEQUENCE	205 AA;	21262 MW;	19FCE67885DFE6A8 CRC64;
 Query Match 36.7%; Score 97.5; DB 1; Length 205; Best Local Similarity 56.9%; Pred.No.0.055; Matches 29; Conservative 4; Mismatches 15; Indels 3; Gaps				
QY	1 AKKYAKKEKAYAKAEKAAKKAAKAYKAAEKKAEEKYKAEAAKAAAKE 51       			
Dd	111 AKVKAK--KAPAKATYAKAAKAAATKA-PAKVAATKAPAKAVKATSPAKK 158       			
 RESULT 12				
TMPB	TREPH			
ID	_TMPB_TREPH	STANDARD;	PRT;	384 AA.
AC	P29720;			
DT	01-APR-1993	(Rel. 25, Created)		
DT	01-APR-1993	(Rel. 25, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Treponema membrane protein B precursor (Antigen tmpB).			
GN	TMPB.			
OS	Treponema phagedenis.			
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.			
ON	NCBI_TaxID=162;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RZ	STRAIN=Kazan 5;			
RC	MEDLINE=91372983;	PubMed=1894368;		
RX	Yelton D.B., Limberger R.J., Curci K., Malinosky-Rummell F.,			
RA	Slivtensky L., Schouls L.M., van Embden J.D., Charon N.W.;			
RT	"Treponema phagedenis encodes and expresses homologs of the Treponema			
RT	pallidum TmpA and TmpB proteins";			
RL	Infect. Immun. 59:3685-3693(1991).			
CC	-I- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR			
CC	LARGE MOLECULES.			
CC	-I- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.			
CC	-I- SIMILARITY: TO TMPB OF T.PALLIDUM.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabor-			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstat-			
CC	the European Bioinformatics Institute. There are no restrictions on co-			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commo-			
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M58563; AAA27480.1; -			
DR	PIR; B43592; B43592.			
KW	Antigen; Outer membrane; Repeat; Signal.			
FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	384	TREPONEMAL MEMBRANE PROTEIN B.
FT	DOMAIN	151	235	17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR]-[ED].
FT	REPEAT	151	155	1-1.
FT	REPEAT	156	160	1-2.
FT	REPEAT	161	165	1-3.
FT	REPEAT	166	170	1-4.
FT	REPEAT	171	175	1-5.
FT	REPEAT	176	180	1-6.
FT	REPEAT	181	185	1-7.
FT	REPEAT	186	190	1-8.
FT	REPEAT	191	195	1-9.
FT	REPEAT	196	200	1-10.
FT	REPEAT	201	205	1-11.
FT	REPEAT	206	210	1-12.
FT	REPEAT	211	215	1-13.
FT	REPEAT	216	220	1-14.
FT	REPEAT	221	225	1-15.
FT	REPEAT	226	230	1-16.
FT	REPEAT	231	235	1-17.
FT	DOMAIN	236	288	6 X 8 AA TANDEM REPEATS OF [EA] -A-A-R-X-

```

FT REPEAT      236      243      A-A-E.
FT REPEAT      245      252      2-1.
FT REPEAT      254      261      2-2.
FT REPEAT      263      270      2-3.
FT REPEAT      272      279      2-4.
FT REPEAT      281      288      2-5.
FT REPEAT      284      291      2-6.
SQ SEQUENCE    384 AA; 42677 MW; 6E94CBC74294DE8C CRC64;

Query Match      36.1%; Score 96; DB 1; Length 384;
Best Local Similarity 51.6%; Pred. No. 0.12;
Matches 33; Conservative 6; Mismatches 17; Indels 8; Gaps 3;

Qy 1 AKKYAKKEKAYAKKA--EKAACK--AEAKAYKAAEAKKA-----EAKYKAAEAKAAKEA 52
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 AKEKAKEKAADKAKEKAKEKAADKAKEKAKEKAAREMAAKEKAADKAADKAKEA 237
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 53 AVEA 56
    ||| |||
Db 238 ARKA 241

RESULT 13
HIG_STRPU
ID_HIG_STRPU STANDARD; PRT; 217 AA.
AC P07796;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-gamma, late.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87172742; PubMed=3031476;
RA Knowles J.A., Lai Z.-C., Childs G.J.;
RT "Isolation, characterization, and expression of the gene encoding the
RT late histone subtype H1-gamma of the sea urchin Strongylocentrotus
RT purpuratus."
RL Mol. Cell. Biol. 7:478-485 (1987).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16033; AAA30059.1; -.
DR PIR; A26721; A26721.
DR HSP; P02259; IHST.
DR InterPro; IPR001386; Histone H1/H5.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Sperm.
FT VARIANT      144      144      K -> R.
SQ SEQUENCE    248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match      35.9%; Score 95.5; DB 1; Length 248;
Best Local Similarity 52.9%; Pred. No. 0.093;
Matches 27; Conservative 4; Mismatches 19; Indels 1; Gaps 1;

Qy 1 AKKYAKKEKAYAKKA--EKAACK--AEAKAYKAAEAKKA--EAKYKAAEAKAAKAAK 50
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 AKAAKKAALAKKAAAKAAAKKAAAKKAAKPKKPKKAAKKAAPKSPKK 195
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
YDF3_SCHPO
ID_YDF3_SCHPO STANDARD; PRT; 1403 AA.
AC Q10475;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable eukaryotic initiation factor C17C9.03.

```

```

Db 135 AKKAATKATKTKKVPAAKKAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 194
Qy 42 AEAATAAAKEA 52
    ||| |||
Db 195 KKAAPKAAKAA 205

RESULT 14
H1_PARAN
ID_H1_PARAN STANDARD; PRT; 248 AA.
AC P02256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal.
OS Parechinus angulosus (Angulate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
OC Parechinus.
OX NCBI_TaxID=7658;
RN [1]
RP SEQUENCE OF 1-84.
RX MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides."
RL Eur. J. Biochem. 104:559-566 (1980).
RN [2]
RP SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure."
RL Eur. J. Biochem. 104:567-578 (1980).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR; A02586; HSURIP.
DR HSP; P02259; IHST.
DR InterPro; IPR001386; Histone H1/H5.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Sperm.
FT VARIANT      144      144      K -> R.
SQ SEQUENCE    248 AA; 26387 MW; 1B25B3F136541947 CRC64;

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 29.654 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYAEAAKAAKEAAEAYEA 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTREMBL 21.\*

2: sp\_archaea.\*

3: sp\_bacteria.\*

4: sp\_fungi.\*

5: sp\_human.\*

6: sp\_invertebrate.\*

7: sp\_mammal.\*

8: sp\_mhc.\*

9: sp\_organelle.\*

10: sp\_phase.\*

11: sp\_plant.\*

12: sp\_rodent.\*

13: sp\_virus.\*

14: sp\_vertebrate.\*

15: sp\_unclassified.\*

16: sp\_rvirus.\*

17: sp\_bacteriap.\*

18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	44.0	372	2 Q9WXX1	Q9wxx1 pseudomonas
2	117	44.0	389	16 Q9CM70	Q9cm70 pasteurella
3	116.5	43.8	239	16 Q8Y5W4	Q8y5w4 listeria mo
4	114.5	43.0	395	2 Q937K4	Q937k4 erwinia chr
5	113.5	42.7	407	16 Q8ZQT6	Q8zqt6 salmonella
6	113	42.5	394	16 Q8X965	Q8x965 escherichia
7	111	41.7	212	3 Q91946	Q91946 candida alb
8	110.5	41.5	376	16 Q8Z8C1	Q8z8c1 salmonella
9	110	41.4	1701	5 Q61164	Q61164 plasmodium
10	104.5	39.3	1866	5 Q8T5C8	Q8t5c8 plasmodium
11	100.5	37.8	200	16 Q8XVW7	Q8xvw7 ralstonia s
12	98.5	37.0	243	16 Q92A67	Q92a67 listeria in
13	98.5	37.0	311	12 Q84528	Q84528 paramecium
14	98	36.8	969	5 Q9ND19	Q9nd19 plasmodium
15	96	36.1	213	3 Q9UV33	Q9uv33 ascobolus i
16	95.5	35.9	629	2 Q9KJ98	Q9kj98 escherichia

17	95	35.7	275	10 Q9XHL9	Q9xhl9 triticum ae
18	95	35.7	388	16 Q8ZGZ2	Q8zgz2 yerzinia pe
19	94.5	35.5	845	5 Q9YIP8	Q9yip8 plasmodium
20	94	35.3	191	5 Q46141	Q46141 mytilus edu
21	94	35.3	191	5 Q46142	Q46142 mytilus edu
22	94	35.3	191	5 Q46143	Q46143 mytilus edu
23	94	35.3	191	5 Q46362	Q46362 mytilus edu
24	94	35.3	703	10 Q9LYA2	Q9lya2 arabidopsis
25	94	35.3	1341	16 Q98KG7	Q98kg7 rhizobium l
26	93	35.0	275	5 Q01395	Q01395 drosophila
27	93	35.0	688	10 Q93VS6	Q93vs6 arabidopsis
28	92.5	34.8	182	2 Q45370	Q45370 bordetella
29	92.5	34.8	1128	5 Q26947	Q26947 trypanosoma
30	92.5	34.8	1671	5 Q8T5C9	Q8t5c9 plasmodium
31	92	34.6	233	5 Q18319	Q18319 chironomus
32	92	34.6	243	2 Q9L564	Q9l564 streptococc
33	92	34.6	247	2 Q9L566	Q9l566 streptococc
34	92	34.6	401	2 Q9LA22	Q9la22 streptococc
35	92	34.6	2055	5 Q8T5C7	Q8t5c7 plasmodium
36	91.5	34.4	236	10 Q9SWU3	Q9swu3 triticum ae
37	91.5	34.4	237	10 Q9SWU2	Q9swu2 triticum ae
38	91.5	34.4	284	10 Q65794	Q65794 triticum ae
39	91	34.2	191	5 Q46140	Q46140 mytilus edu
40	90.5	34.0	581	16 Q9EU45	Q9eu45 deinococcus
41	90.5	34.0	755	10 Q9FP71	Q9fp71 oryza sativ
42	90	33.8	101	5 Q9BMY8	Q9bmy8 leishmania
43	90	33.8	345	16 Q8Y1F6	Q8y1f6 ralstonia s
44	89.5	33.6	136	10 Q39681	Q39681 daucus caro
45	89.5	33.6	356	16 Q9KR10	Q9kr10 vibrio chol

#### ALIGNMENTS

#### RESULT 1

ID	Q9WXX1	PRELIMINARY;	PRT;	372 AA.
AC	Q9WXX1;			
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)			
DE	Tola protein.			
GN	TOLA.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=303;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RX	MEDLINE=96198174; PubMed=8626299;			
RA	Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;			
RT	"the Pseudomonas putida peptidoglycan-associated outer membrane			
RT	lipoprotein (PAL) is involved in maintenance of the integrity of the			
RT	cell envelope.";			
RL	J. Bacteriol. 178:1699-1706(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RA	Ramos-Gonzalez I.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RA	Rodriguez-Herva J.J.;			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RX	MEDLINE=96422022; PubMed=8824639;			
RA	Rodriguez-Herva J.J., Ramos J.;			
RT	"Characterization of an OprL null mutant of Pseudomonas putida.";			
RL	J. Bacteriol. 178:5836-5840(1996).			









```

RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RD EMBL; U42580; AAC96576.1; -
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
SQ SEQUENCE 311 AA; 35390 MW; 97B0E9E82AFEF88A CRC64;

Query Match 37.0%; Score 98.5; DB 12; Length 311;
Best Local Similarity 43.9%; Pred.No. 0.13;
Matches 29; Conservative 8; Mismatches 18; Indels 11; Gaps

Qy 2 KYAKKEKAYAK-----KAEEAAKAAKAEKAYKAAEAKK-----KAEEKYKAEAAKAAK 50
Db | : | | | | : ||||| : | | | | : | | | | : | | | | : | | | | : | | | | :
72 KERVVKAARAEAEKERVVKAERIKAEKAVEKERIKAEKAEKRVVKAEPAPAR 131

Qy 51 EAAYEA 56
Db | | | | |
132 EAAKAA 137

RESULT 14
Q9NDI9 PRELIMINARY; PRT; 969 AA.
AC Q9NDI9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DD 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 3g.
DN MSP-3G.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BELEM;
RX MEDLINE=21273153; PubMed=11377738;
RA Galinski M.R., Ingravallo P., Corredor-Medina C., Al-Khedery B.,
RA Povoa M., Barnwell J.W.;
RT "Plasmodium vivax merozoite surface proteins-3beta and-3gamma share structural similarities with P. vivax merozoite surface protein-3alpha and define a new gene family.";
RL Mol. Biochem. Parasitol. 115:41-53(2001).
RD EMBL; AF099663; AAF78288.1; -.
SQ SEQUENCE 969 AA; 104674 MW; 9A150CCA9918FF77 CRC64;

Query Match 36.8%; Score 98; DB 5; Length 969;
Best Local Similarity 39.4%; Pred.No. 0.44;
Matches 26; Conservative 11; Mismatches 17; Indels 12; Gaps

Qy 1 AKKYAKKEKAYAKKAEEAAKAEKAYKAAEKKAEEAK-----YKAEAAKAA 48
Db | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
594 AKQAEEKAEKITTKATEANKAKEBEKAEKAEKAEKAGDVEDVEVAVNVFESVKAA 653

Qy 49 KEAAAY 54
Db | | | | |
654 AKAAAH 659

RESULT 15
Q9UV33 PRELIMINARY; PRT; 213 AA.
ID Q9UV33
AC AC
AC Q9UV33;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DD 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Histone H1.
DN H1.
OS Asacbolus immersus.
```



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	266	100.0	56	21	AA92573	Copolymer molecule	
2	245.5	92.3	77	21	AA92575	Copolymer molecule	
3	233	87.6	86	21	AA92576	Copolymer molecule	
4	198	74.4	66	21	AA92577	Copolymer molecule	
5	180.5	67.9	109	21	AA92577	Copolymer molecule	
6	134.5	50.6	45	21	AA92572	Copolymer molecule	
7	116.5	43.8	239	23	AB949123	Listeria monocytog	
8	111	41.7	106	11	AA96446	Recombinant copoly	
9	104.5	39.3	35	21	AA92571	Copolymer molecule	
10	104	39.1	100	21	AA98499	Peptide #10 used i	

PA (TEVA-) TEVA PHARM USA INC.





AAY82577  
 ID AAY82577 standard; peptide; 109 AA.  
 XX  
 XX AAY82577;  
 AC  
 XX  
 DT 28-JUL-2000 (first entry)  
 XX  
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ\_ID NO:7.  
 XX  
 KW Copolymer; molecular weight marker; TV-marker; immune disease;  
 KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thymostimetic; haemostatic; antipsoriatic; dermatological;  
 KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
 KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.  
 XX  
 OS Unidentified.  
 XX  
 PN W0200018794-A1.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 24-SEP-1999; 99WO-US22402.  
 XX  
 PF 25-SEP-1998; 98US-0101693.  
 XX  
 PR (YEDA ) YEDA RES & DEV CO LTD.  
 PA (TEVA-) TEVA PHARM USA INC.  
 PA  
 XX  
 PI Gad A, Lis D;  
 XX  
 DR WPI; 2000-317499/27.  
 XX  
 XX  
 PT Copolymer 1 related polypeptides used as molecular weight markers for  
 PT glatiramer acetate and for treatment and prevention of immune diseases  
 XX  
 PS Claim 10; Page 14; 72pp; English.  
 XX  
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular  
 CC weight of a copolymer (CP), which has an identified molecular weight  
 CC and an amino acid composition corresponding to the copolymer. The  
 CC polypeptides of the invention are used as molecular weight markers for  
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
 CC used for treating and preventing immune diseases in a mammal. Autoimmune  
 CC diseases which may be treated include either cell-mediated or  
 CC antibody-mediated diseases. Such diseases include arthritic conditions, or  
 CC demyelinating diseases and inflammatory conditions, e.g. multiple  
 CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 CC lupus erythematosus. Mediated-mediated diseases which can be treated  
 CC include host-versus-graft disease, graft-versus-host disease, and  
 CC delayed-type hypersensitivity. The polypeptides of the invention have  
 CC defined molecular weights and physical properties which are analogous to  
 CC glatiramer acetate molecules, which makes them ideal for use as  
 CC molecular weight markers.  
 XX  
 SQ Sequence 109 AA;  
 Query Match 67.9%; Score 180.5; DB 21; Length 109;  
 Best Local Similarity 45.9%; Pred. No. 2.6e-11;  
 Matches 50; Conservative 3; Mismatches 3; Indels 53; Gaps 3  
 QY 1 AKYAKK-ERAYAKCA-----EKAAKKAKAYKAAEKKAQK----- 39

[illegible]











Search completed: March 10, 2003, 12:21:29  
Job time : 30.1814 secs

THIS PAGE BLANK (USPTO)



```
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match          92.3%; Score 245.5; DB 10; Length 77;
Best Local Similarity 72.7%; Pred. No. 5.3e-18;
Matches 56; Conservative 0; Mismatches 0; Indels 21; Gaps 1;

Qy 1 AKYAKKEKAYAKAEKAAKAEKAYKAAAEAKKK-----AEAK 39
    |||||
Db 1 AKYAKKEKAYAKAEKAAKAEKAYKAAAEKKKAKAEAKKYAKAAKAEKKEKYEAAAEAK 60
    |||||

Qy 40 YKAEAAKAAAEKAEAYEA 56
    |||||
Db 61 YKAEAAKAAAEKAEAYEA 77

RESULT 3
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match          87.6%; Score 233; DB 10; Length 86;
Best Local Similarity 62.8%; Pred. No. 9.9e-17;
Matches 54; Conservative 2; Mismatches 0; Indels 30; Gaps 1;

Qy 1 AKYAKKEKAYAKAEKAAKAEKAYKAAAEAKKKAEAK----- 39
    |||||
Db 1 AKYAKKEKAYAKAEKAAKAEKAYKAAAEKKKAKAEAKKYAKAAKAEKKEKYEAAAEAK 60
    |||||

Qy 40 -----YKAEAAKAAAEKAEAYEA 56
    |||||
Db 61 YKAEAAKAYKAEKAAKAAAEKAEAYEA 86

RESULT 4
US-09-816-989A-4
```

```
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

Query Match          74.4%; Score 198; DB 10; Length 66;
Best Local Similarity 80.3%; Pred. No. 2e-13;
Matches 53; Conservative 2; Mismatches 1; Indels 10; Gaps 5;

Qy 1 AKKYAKKEKAY--AKAE--KAAK--KAEAKAY-KAAEAKKK-----AEAKYKAAAEKAAAK 50
    |||||
Db 1 AKKYAKKEKAYAKAKAEKAAKAEKAYKAAAEKKEKYEAAAEAKYKAAAEKAAAK 60
    |||||

Qy 51 EAAEYEA 56
    |||||
Db 61 EAAEYEA 66

RESULT 5
US-09-816-989A-7
; Sequence 7, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match          67.9%; Score 180.5; DB 10; Length 109;
Best Local Similarity 45.9%; Pred. No. 1.7e-11;
Matches 50; Conservative 3; Mismatches 3; Indels 53; Gaps 3;

Qy 1 AKKYAKK-EKAYAKKA-----EKAACKAEAKYKAAEAKKKAEAK----- 39
    |||||
Db 1 AKKYAKKEKAYAKAKAEKKEKAYAKAEKKEKYEAAAEKKEKYEAAAEKKEA 60
    |||||

Qy 40 -----YKAEAAKAAAEKAEAYEA 56
    |||||
```



Db 61 YKAEAKYAKAAKAEKKEYYAAAEAKKAEAKAYKAEAAKAAKAAEAYEA 109

## RESULT 6

US-09-816-989A-2  
; Sequence 2, Application US/09816989A  
; Patent No. US20020115103A1  
; GENERAL INFORMATION:  
; APPLICANT: Gad, Alexander  
; APPLICANT: Lis, Doris  
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK  
; FILE REFERENCE: 2609/60807-A-PCT-US  
; CURRENT APPLICATION NUMBER: US/09/816,989A  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 45  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide  
US-09-816-989A-2

Query Match 50.6%; Score 134.5; DB 10; Length 45;  
Best Local Similarity 68.4%; Pred. No. 2.3e-07;  
Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

Qy 1 AKYAKKEKAYAKKAEKAAKAEKAYKAEAKKAEAKY-KAEAAKAAKAAEAYEA 56  
Db 1 AKYAKK-----AKAEKA-----KKAYKAEAKK--AAKYKAEAAKAEKAAEAYEA 45

## RESULT 7

US-09-820-843A-8  
; Sequence 8, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; FILE REFERENCE: 063915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: H. influenzae  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: outer membrane integrity protein (tolA)  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|1573353  
US-09-820-843A-8

Query Match 39.7%; Score 105.5; DB 9; Length 372;  
Best Local Similarity 48.2%; Pred. No. 0.0013;  
Matches 27; Conservative 14; Mismatches 14; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKKAEKAAK-KAEAKAYKAEAKKAEAKYKAEAAKAAKAAEAYEA 56  
Db 133 EKQKQAEAKQAEEAKKLADAEAKRLAAAKQAEEAKAKAAETAAQKAKQEA 188

## RESULT 8

US-09-816-989A-1

; Sequence 1, Application US/09816989A  
; Patent No. US20020115103A1  
; GENERAL INFORMATION:  
; APPLICANT: Gad, Alexander  
; APPLICANT: Lis, Doris  
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK  
; FILE REFERENCE: 2609/60807-A-PCT-US  
; CURRENT APPLICATION NUMBER: US/09/816,989A  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide  
US-09-816-989A-1

Query Match 39.3%; Score 104.5; DB 10; Length 35;  
Best Local Similarity 51.8%; Pred. No. 0.00015;  
Matches 29; Conservative 4; Mismatches 2; Indels 21; Gaps 2;

Qy 1 AKYAKKEKAYAKKAEKAAKAEKAYKAEAKKAEAKYKAEAAKAAKAAEAYEA 56  
Db 1 AKYAKKEKA-----AKYAKK-----EAKAKAAEAAKAEAYEA 35

## RESULT 9

US-10-051-643-201  
; Sequence 201, Application US/10051643  
; Publication No. US20020197265A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Tan, Paul L. J.  
; TITLE OF INVENTION: Methods and Compounds for the Treatment  
; TITLE OF INVENTION: Of Immunologically-Mediated Diseases of the Respiratory  
; FILE REFERENCE: 11000.1008G2  
; CURRENT APPLICATION NUMBER: US/10/051,643  
; CURRENT FILING DATE: 2002-01-18  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR FILING DATE: 1997-12-23  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 201  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Mycobacterium vaccae  
US-10-051-643-201

Query Match 38.7%; Score 103; DB 9; Length 223;  
Best Local Similarity 55.0%; Pred. No. 0.0014;  
Matches 33; Conservative 3; Mismatches 18; Indels 6; Gaps 2;

Qy 1 AKYAKKEKAYAKKAEKAAKAEAK-----AYKAAEAKKKAEAKYKAEAAKAAKAAEAYEA 56  
Db 126 AKKAAPAKKAPAKKA--ATKAAPAKYATAAKKAPAKKATAAKKAPAKKAPAKKATKA 183

## RESULT 10

US-10-184-832-5  
; Sequence 5, Application US/10184832  
; Publication No. US20030022857A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu et al.





THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 9.92405 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYKAEAKAAKAEAYEA 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	39.1	56	4	US-08-993-008A-6
2	104	39.1	100	2	US-08-460-890A-64
3	104	39.1	100	3	US-08-167-641C-64
4	104	39.1	100	4	US-08-460-371A-64
5	104	39.1	100	4	US-08-462-040-64
6	103	38.7	223	4	US-09-095-855-201
7	103	38.7	223	4	US-09-205-426-201
8	101	38.0	1507	3	US-08-929-329-5
9	99.5	37.4	48	4	US-08-993-008A-5
10	94.5	35.5	214	3	US-09-041-889-27
11	94	35.3	433	1	US-08-346-849-2
12	94	35.3	433	2	US-08-293-284A-2
13	92.5	34.8	643	2	US-08-216-894-8
14	92.5	34.8	643	4	US-09-115-746-8
15	91.5	34.4	60	1	US-08-346-849-16
16	91.5	34.4	60	2	US-08-293-284A-16
17	91.5	34.4	472	2	US-08-216-894-10
18	91.5	34.4	472	4	US-09-115-746-10
19	91.5	34.4	564	2	US-08-216-894-2
20	91.5	34.4	564	4	US-09-115-746-2
21	90.5	34.0	1507	6	5268270-2
22	84	31.6	32	1	US-08-152-488-13
23	84	31.6	32	1	US-08-303-025-15
24	84	31.6	32	1	US-08-677-304-13
25	84	31.6	32	2	US-08-436-703B-2
26	84	31.6	33	1	US-08-303-025-16
27	84	31.6	33	2	US-08-436-703B-4

28	82	30.8	288	3	US-08-312-949-4	Sequence 4, Appli
29	82	30.8	288	3	US-08-446-201-4	Sequence 4, Appli
30	82	30.8	289	1	US-08-072-070-4	Sequence 4, Appli
31	82	30.8	289	1	US-08-469-434-4	Sequence 4, Appli
32	82	30.8	289	1	US-08-214-222-4	Sequence 4, Appli
33	82	30.8	289	2	US-08-467-852A-5	Sequence 5, Appli
34	82	30.8	289	2	US-08-468-718-4	Sequence 4, Appli
35	82	30.8	289	2	US-08-247-491A-5	Sequence 5, Appli
36	82	30.8	619	1	US-08-465-746-2	Sequence 2, Appli
37	82	30.8	619	1	US-08-214-164-2	Sequence 2, Appli
38	82	30.8	619	2	US-08-467-852A-3	Sequence 3, Appli
39	82	30.8	619	2	US-08-246-636-2	Sequence 2, Appli
40	82	30.8	619	2	US-08-247-491A-3	Sequence 3, Appli
41	82	30.8	619	2	US-08-319-795-2	Sequence 2, Appli
42	82	30.8	619	2	US-08-468-985-2	Sequence 2, Appli
43	82	30.8	619	3	US-08-312-949-2	Sequence 2, Appli
44	82	30.8	648	1	US-08-072-070-2	Sequence 2, Appli
45	82	30.8	648	1	US-08-469-434-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-993-008A-6  
; Sequence 6, Application US/08993008A  
; Patent No. 6153596  
; GENERAL INFORMATION:  
; APPLICANT: Liotta, Dennis C.  
; APPLICANT: Petros, John A.  
; APPLICANT: Wey, Shioh-Jyi  
; APPLICANT: Karr, Joan F.  
; APPLICANT: Pohl, Jan  
; TITLE OF INVENTION: Polycationic Oligomers  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/993,008A  
; FILING DATE: 18-DEC-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/032,436  
; FILING DATE: 18-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sullivan, Sally A.  
; REGISTRATION NUMBER: 32,064  
; REFERENCE/DOCKET NUMBER: 33-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303-499-8080  
; TELEFAX: 303-499-8089  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 56 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-993-008A-6

Query Match 39.1%; Score 104; DB 4; Length 56;











RESULT 13  
 US-08-216-894-8  
 ; Sequence 8, Application US/08216894  
 ; Patent No. 5876734  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kirchhoff, Louis V.  
 ; APPLICANT: Otsu, Keiko  
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION

[illegible]

NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-8

Query Match 34.8%; Score 92.5; DB 4; Length 643;  
Best Local Similarity 47.5%; Pred. No. 0.039;  
Matches 29; Conservative 6; Mismatches 15; Indels 11; Gaps 2;

Qy 1 AKYAKKEKAYAKAEKAAKAEKAYKAEAKKAEAKYK-----AEAKAAAEK 52  
Db 504 ATKVAEEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 560

Qy 53 A 53  
Db 561 A 561

RESULT 15  
US-08-346-849-16  
Sequence 16, Application US/08346849  
Patent No. 5670483  
GENERAL INFORMATION:  
APPLICANT: Zhang, Shuguang  
APPLICANT: Lockshin, Curtis  
APPLICANT: Rich, Alexander  
APPLICANT: Holmes, Todd  
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY  
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES  
TITLE OF INVENTION: THEREFOR  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/346,849  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,326  
FILING DATE: 28 DECEMBER 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: MIT-6008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-346-849-16

Query Match 34.4%; Score 91.5; DB 1; Length 60;  
Best Local Similarity 52.0%; Pred. No. 0.0041;  
Matches 26; Conservative 4; Mismatches 19; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAEKAAK-KAEAKAYKAAEAKKAEAKYKAEAKAAK 50  
Db 1 KAAAKRKAALAKKAAKAAKRAAAKAAKAKKAKKKKAAKCAKPKAKSPKK 50

Search completed: March 10, 2003, 12:30:02  
Job time : 9.92405 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 19.3544 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989A-4  
Perfect score: 313  
Sequence: 1 AKYAKKEKAYAKAKAEAK.....EAKYAKBAKAAKAAKAAAYEA 66

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.5	42.3	394	2 F90725	membrane spanning
2	132.5	42.3	394	2 G85576	membrane spanning
3	131	41.9	421	2 JV0057	tola protein - Bsc
4	123	39.3	248	1 HSUR1P	histone H1, gonada
5	122.5	39.1	376	2 AG0592	tola protein (impo
6	121.5	38.8	1390	2 S51364	sperm tail-specifi
7	119.5	38.2	347	2 E83525	TOLA protein PA097
8	118	37.7	231	2 S59589	histone H1 - Chlam
9	117.5	37.5	372	2 G84064	outer membrane int
10	117	37.4	388	2 AC0138	TOLA colicin impor
11	115.5	36.9	206	2 S09388	histone H1 - sea u
12	115	36.7	344	2 S34153	mst101-1 protein -
13	113	36.1	284	2 T06241	histone H1 (clone
14	112.5	35.9	217	2 A26721	histone H1-gamma,
15	112	35.8	243	2 A21689	hypothetical prote
16	111.5	35.6	239	2 A21317	hypothetical prote
17	111	35.5	210	2 A25550	histone H1 - sea u
18	110	35.1	220	2 A28456	histone H1.10 - ch
19	109	34.8	209	1 HSKLIA	histone H1A - Afri
20	107	34.2	384	2 B43592	outer membrane pro
21	107	34.2	1701	2 T09127	probable erythrocy
22	106	33.9	214	2 G70673	probable hubp - My
23	105	33.5	211	2 A28100	histone H1-beta, e
24	104	33.2	311	2 T17698	hypothetical prote
25	103	32.9	182	2 S61926	histone H1 homolog
26	102.5	32.7	433	2 S25194	zuotin - yeast (Sa
27	102	32.6	924	2 T06636	hypothetical prote
28	101.5	32.4	236	2 S22322	histone H1 - wheat
29	101.5	32.4	1128	2 T30296	R27-2 protein - Tr

30 101 32.3 206 1 HSTR1R histone H1 - rainb  
31 101 32.3 829 2 E64114 translation initia  
32 99.5 31.8 229 2 I51227 histone H1A - Afri  
33 99 31.6 225 2 B28456 histone H1.11L - c  
34 99 31.6 581 2 E75383 conserved hypochet  
35 98.5 31.5 218 2 S01262 histone H1 - musco  
36 98.5 31.5 241 2 JN0748 histone H1-II - Vo  
37 98.5 31.5 328 2 A44993 cytosolic repetiti  
38 98 31.3 288 2 T06257 histone H1 (clone  
39 97.5 31.2 356 2 A82152 toIA protein VC183  
40 97.5 31.2 438 2 G87675 arylesterase-relat  
41 97 31.0 224 2 D28456 histone H1.03 - ch  
42 96.5 30.8 194 1 HSTR1 histone H1 - trout  
43 96.5 30.8 219 1 HSL1B histone H1B - Afri  
44 96.5 30.8 220 2 I51447 histone H1B - Afri  
45 96.5 30.8 221 2 S33219 histone H1.C - Afr

## ALIGNMENTS

## RESULT 1

F90725

membrane spanning protein Tola [imported]

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: F90725

R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: F90725

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34197.1; PID:gl3360233; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs0774

Query Match 42.3%; Score 132.5; DB 2; Length 394;  
Best Local Similarity 58.9%; Pred. No. 0.00099;  
Matches 43; Conservative 8; Mismatches 13; Indels 9; Gaps 4;

Qy 1 AKYAKKEKAYAKAKAEAKKAEAKKYAKAA-----KAEKKEYAAAEAKYKAEA 54

Db 156 AEEAAKKAADAK-KKAEAEA-KAAEAQKKAEEAAALKKKAEAAAEAAAEARKKAAA 213

Qy 55 AKAAA-KEAAVEA 66

Db 214 EKAADKKAEEKA 226

## RESULT 2

G85576

membrane spanning protein Tola [imported]

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Nov-2001

C:Accession: G85576

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85576

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <STO>

A:Cross-references: GB:AE005174; NID:gl2513672; PIDN:AAG55075.1; GSPDB:GN00145; UWGP:Z09

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: toIA









A;Reference number: A26721; MUID:87172742; PMID:3031476

A;Accession: A26721

A;Molecule type: DNA

A;Residues: 1-217 <KNO>

A;Cross-references: GB:M16033; NID:g161517; PIDN:AAA30059.1; PID:g161518

C;Superfamily: histone H1

C;Keywords: Chromosomal protein; DNA binding; nucleosome; nucleus

F;2-217/Product: histone H1-gamma, embryonic #status predicted <NAT>

Query Match 35.9%; Score 112.5; DB 2; Length 217;  
Best Local Similarity 49.4%; Pred. No. 0.023;  
Matches 38; Conservative 3; Mismatches 21; Indels 15; Gaps 3;

Qy 1 AKKYAKKEK-----AYAKAKAEAK-----AAKKAKAEAKKYAKAAK---AEKKEYAA 45

Db 139 AKKATPKTKVKPKPAKKPAKKPAKKPAKKPAKKPAKKPAKKPAKKPAKKPAKKPAKKPAKKAA 198

Qy 46 AEAKYKAEAAKAAKEA 62

Db 199 KPAKKAAPKAAKPA 215

#### RESULT 15

AE1689

hypothetical protein homolog lin2055 [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AE1689

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.

A;Reference number: AE1077; MUID:21537279; PMID:11679669

A;Accession: AE1689

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-243 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC97285.1; PID:g16414556; GSPDB:GN00178

A;Experimental source: strain Clip11262

C;Genetics:

A;Gene: lin2055

Query Match 35.8%; Score 112; DB 2; Length 243;  
Best Local Similarity 47.1%; Pred. No. 0.028;  
Matches 32; Conservative 11; Mismatches 23; Indels 2; Gaps 1;

Qy 1 AKKYAKKEKAYAKAKAEAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 58

Db 110 ABEKAAAEKAAEBEKAADKKSQDEAAKAAAKKQEAEEKAAAEKAAADKAA 169

Qy 59 AKEAAYEA 66

Db 170 KEKAAEA 177

Search completed: March 10, 2003, 12:28:27

Job time : 19.3544 secs

**THIS PAGE BLANK (USPTO)**

RESULT 1
ID TOLA ECOLI
STANDARD; PRT; 421 AA.
P19934;
AC 01-FEB-1991 (Rel. 17, Created)
DT DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
GN TOLA OR CIM OR EXCC OR LKY OR B0739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision;
OC Escherichia.
OX NCBI_Taxid=562;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=JM105;
RC MEDLINE=90078104; PubMed=2687247;
RX Levensgood S.K., Webster R.E.;
RA "Nucleotide sequences of the <i>tola</i> and <i>tolB</i> genes
RT their products, components of a multistep trans-
RL Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RX Blattner F.R.; Plunkett G. III, Bloch C.A., Fanning
RA Riley M., Collado-Vides J., Glasner J.D., Rood
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goetz
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia
RL Science 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RC MEDLINE=97061202; PubMed=8905232;
RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kimura
RA Kimura S., Kitagawa M., Makino K., Masuda S.,
Mori H., Motomura K., Nakamura Y., Nishimoto
RA Sampei G., Seki Y., Tagami H., Takemoto K., Yanai
RA Yano M., Horiiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli
RT corresponding to the 12.7-28.0 min region on
RL DNA Res. 3:137-155(1996).
[4]
RN DOMAINS.
RP MEDLINE=91296736; PubMed=2068069;
RX Levensgood S.K., Beyer W.F. Jr., Webster R.E. E.
RA "Tola: a membrane protein involved in colicin
RT extended helical region."
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991)
[5]
RN INTERACTION WITH PORINS.
RP MEDLINE=97133271; PubMed=8978668;
RX Derouiche R., Gavioli M., Benedetti H., Prilipov

RA Llobes R.;  
RT "TolA central domain interacts with Escherichia coli porins.";  
RL EMBO J. 15:6408-6415(1996).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS) OF 298-421.  
RX MEDLINE=99332679; PubMed=10404600;  
RA Lubkowski J., Henneke F., Plueckthun A., Wlodawer A.;  
RT "Filamentous phage infection: crystal structure of g3p in complex  
RT with its coreceptor, the C-terminal domain of TolA.";  
RL Structure 7:711-722(1999).  
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A  
CC COICINS TO REACH A, EI, E2, E3, AND K). NECESSARY FOR THE  
CC COICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL  
CC BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION  
CC OF BACTERIOPHAGE DNA.  
CC -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, PHOE  
CC AND LAMB.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M28232; AAA24683.1; -;  
DR EMBL; AE000177; AAC73833.1; -;  
DR EMBL; D90713; BAA35405.1; -;  
DR PIR; JY0057; JV0057.  
DR PDB; 1TOL; 20-MAY-99.  
DR EcoGene; EG11007; tola.  
KW Transport; Protein transport; Bacteriocin transport; Transmembrane;  
KW Repeat; Inner membrane; 3D-structure; Complete proteome.  
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 14 34 POTENTIAL.  
FT DOMAIN 35 421 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).  
FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).  
FT DOMAIN 224 278 10 X TANDEM REPEATS OF [ED]-K(1,2)-  
FT A(2,4).  
SQ SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;

Query Match 41.9%; Score 131; DB 1; Length 421;  
Best Local Similarity 52.6%; Pred. No. 0.00058;  
Matches 41; Conservative 9; Mismatches 16; Indels 12; Gaps 3;  
Qy 1 AKYAKKEKAYAKAK-----KAEAKAKAKAEAKKYAK-----AAKAEKK-EYAAAEA 48  
Db 134 AEEAAKAAADAKAKAEADAKAEAEAKKAAADAKKAEAEAKAEAKAEAKKAEAAAL 193  
Qy 49 KYKAEAKAAKAEAYEA 66  
Db 194 KKKAEAEAEAEAEAKKA 211

RESULT 2  
HL\_PARAN  
ID HI\_PARAN STANDARD; PRT; 248 AA.  
AC P02256;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Histone H1, gonsdail.  
OS Parechinus angulosus (Angulate sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;  
OC Parechinus.  
OX NCBI\_TaxID=7658;  
RN [1]  
RP SEQUENCE OF 1-84.  
RX MEDLINE=80156831; PubMed=6767609;

RA Strickland W.N., Strickland M., de Groot P.C., von Holt C.,  
RA Wittmann-Liebold B.;  
RT "The primary structure of histone H1 from sperm of the sea urchin  
RT Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the  
RT protein and the sequence of amino acids in the four N-terminal  
RT cyanogen bromide peptides.";  
RL Eur. J. Biochem. 104:559-566(1980).  
RN [2]  
RP SEQUENCE OF 80-248.  
RX MEDLINE=80156832; PubMed=7363905;  
RA Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,  
RA Wittmann-Liebold B.;  
RT "The primary structure of histone H1 from sperm of the sea urchin  
RT Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and  
RT the entire primary structure.";  
RL Eur. J. Biochem. 104:567-578(1980).  
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF  
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.  
CC -!- SUBCELLULAR LOCATION: Nucleur.  
CC -!- TISSUE SPECIFICITY: SPERM.  
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.  
DR PIR; A02586; HSUR1P.  
DR HSP; P02259; LHST.  
DR InterPro; IPR001386; Histone H1/H5.  
DR InterPro; IPR003216; Linkerhist N.  
DR Pfam; PF00538; linker histone; l.  
DR ProDom; PD000373; Linkerhist\_N; 1.  
DR SMART; SM00526; H15; 1.  
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;  
KW Sperm.  
FT VARIANT 144 144 K -> R.  
SQ SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match 39.3%; Score 123; DB 1; Length 248;  
Best Local Similarity 56.1%; Pred. No. 0.0016;  
Matches 37; Conservative 7; Mismatches 20; Indels 2; Gaps 2;  
Qy 3 KYAKKEKAYAKAEAKAKKA-KAEAKKYAKAEAKKEKAA-AEAKYKAEAKAAK 60  
Db 120 KKAKTSAAKAKAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 179  
Qy 61 EAYEA 66  
Db 180 KAAKKA 185

RESULT 3  
MST2\_DROHY  
ID MST2\_DROHY STANDARD; PRT; 1391 AA.  
AC Q08696;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Axoneme-associated protein mst101(2).  
GN MST101(2).  
OS Drosophila hydei (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7224;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=95045538; PubMed=7957199;  
RA Neesen J., Padmanabhan S., Buenemann H.;  
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid  
RT motif representing the major component of the sperm-tail-specific  
RT axoneme-associated protein family Dhmst101 form extended  
RT alpha-helical rods within the extremely elongated spermatozoa of  
RT Drosophila hydei.";  
RL Eur. J. Biochem. 225:1089-1095(1994).  
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY







DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).  
 GN HUP OR HLP OR MDP1.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ANS;  
 RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
 RT "HUPmt-A target for differentiation of M.tuberculosis and M.bovis.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BCG / Tokyo;  
 RA Magunato S., Yukitake H., Matsuo T., Minoda T., Yamada T.;  
 RT "Identification of a novel protein generating bacterial slow growth  
 from Mycobacterium";  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
 CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
 CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
 CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y18421; CAB46493.1; -;  
 DR EMBL; AB013441; BAA78330.1; -;  
 DR HSSP; P02346; 1HUU.  
 DR InterPro; IPR000119; Bac DNABind.  
 DR InterPro; IPR001386; Histone\_H1/H5.  
 DR Pfam; PF00216; Bac DNA binding; 1.  
 DR PRINTS; PD00624; HISTONERS.  
 DR ProDom; PD000945; Bac DNABind; 1.  
 DR SMART; SM00411; BHL; 1.  
 DR PROSITE; PS00045; HISTONE LIKE; 1.  
 DR PROSITE; PS00045; HISTONE LIKE; Repeat.  
 KW DNA-binding; DNA condensation; Repeat.  
 FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
 FT DOMAIN 101 205 DEGENERATE REPEATS REGION.  
 FT CONFLICT 199 199 A -> T (IN REF. 2).  
 SQ SEQUENCE 205 AA; 21262 MW; 19FCE67885DFE6A8 CRC64;  
 Query Match 35.3%; Score 110.5; DB 1; Length 205;  
 Best Local Similarity 47.0%; Pred. No. 0.013;  
 Matches 39; Conservative 5; Mismatches 22; Indels 17; Gaps 4;  
 Qy 1 AKYAKK---EKAYAKAKAEAKA-AKKA--KAEAKYAKAKAEKKEKAYAAA-----46  
 Db 111 AKVAKKAPAKKATKAAKAATKAPAKKAATKAPAKKAVKATKSPAKKVTKVAKTKAVKA 170  
 Qy 47 ---EAKYAKAEAKAKAEKAEVEA 66  
 Db 171 SVRKAATKAPAKKAAKRRPATKA 193  
 RESULT 11  
 ID H10\_CHICK STANDARD; PRT; 219 AA.  
 AC P08286;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Histone H1.10 (Methylated DNA binding protein-2-H1) (MDBP-2-H1).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87250632; PubMed=3597432;  
 RA Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;  
 RT "Characterization of the chicken histone H1 gene complement.  
 Generation of a complete set of vertebrate H1 protein sequences.";  
 RL J. Biol. Chem. 262:9656-9663(1987).  
 RN [2]  
 RP IDENTIFICATION OF ALA-13.  
 RX MEDLINE=98060905; PubMed=9396815;  
 RA Schwarz S., Hess D., Jost J.P.;  
 RT "The methylated DNA binding protein-2-H1 (MDBP-2-H1) consists of  
 RT histone H1 subtypes which are truncated at the C-terminus";  
 RL Nucleic Acids Res. 25:5052-5056(1997).  
 CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF  
 CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M17018; AAA48788.1; -;  
 DR PIR; A28456; A28456.  
 DR HSSP; P08287; 1GHC.  
 DR InterPro; IPR001386; Histone\_H1/H5.  
 DR Pfam; PF00538; linker\_histone; 1.  
 DR ProDom; PD000373; Linkerhist\_N; 1.  
 DR SMART; SM00526; H15; 1.  
 DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;  
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;  
 FT INIT\_MET 0 0  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT DOMAIN 36 109 GLOBULAR.  
 SQ SEQUENCE 219 AA; 21872 MW; A94DF6C0D3048AE CRC64;  
 Query Match 35.1%; Score 110; DB 1; Length 219;  
 Best Local Similarity 48.4%; Pred. No. 0.015;  
 Matches 30; Conservative 5; Mismatches 27; Indels 0; Gaps 0;  
 Qy 2 KKYAKKAYAKAKAEAKAKAEAKAKAEAKAKAEAKAKAEAKAEAKAKAEAKAEAKAE 61  
 Db 144 KCAAAAKSPKAKKAPAAATTKAAKSPKATKAAPKKAATKSPAKAKAVKPKAAKPK 203  
 Qy 62 AA 63  
 Db 204 AA 205  
 RESULT 12  
 ID H1A\_XENLA STANDARD; PRT; 209 AA.  
 AC P06892;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Histone H1A.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A. (GENE CLUSTER X1H3).  
 RX MEDLINE=86037224; PubMed=3863963;





RP SEQUENCE FROM N.A.  
RC STRAIN=sv. Horoshirikomugi; TISSUE=Seedling;  
RX MEDLINE=92020152; PubMed=1923777;  
RA Yang P., Katsura M., Nakayama T., Mikami K., Iwabuchi M.;  
RT "Molecular cloning and nucleotide sequences of cDNAs for histone H1  
and H2B variants from wheat.";  
RL Nucleic Acids Res. 19:5077-5077(1991).  
RN [2]  
RP REVISIONS.  
RC STRAIN=sv. Horoshirikomugi;  
RA Morisawa G., Meshi T., Iwabuchi M.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF  
NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X59872; CAA42529.2; -.  
DR PIR; S22322; S22322.  
DR HSSP; P02259; LHST.  
DR InterPro; IPR001386; Histone\_H1/H5.  
DR InterPro; IPR0013216; Linkerhist\_N.  
DR Pfam; PF00538; linker histone; 1.  
DR PRINTS; PR00624; HISTONEHS.  
DR ProDom; PD000373; Linkerhist\_N; 1.  
DR SMART; SM00526; H15; 1.  
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.  
SQ SEQUENCE 238 AA; 24222 MW; 2CBIC1699B65A27 CRC64;  
-----  
Query Match 34.0%; Score 106.5; DB 1; Length 238;  
Best Local Similarity 56.2%; Pred. No. 0.03;  
Matches 36; Conservative 5; Mismatches 20; Indels 3; Gaps 3;  
QY 2 KYAKKEKAYAKAKAEAKA-AKKAKEAKKYAKA-AKAEKKEAYAAEKYAE-AKAA 58  
Db 137 KKPAKPAKPAKPAKTAAPKAAKPAKAAKPAKAAKPAKAAKPAKAAKPAKAA 196  
QY 59 AKEA 62  
Db 197 AKKA 200  
-----  
RESULT 15  
DBH\_MYCTU  
ID DBH\_MYCTU STANDARD; PRT; 214 AA.  
AC P95109;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp) (21-kDa  
laminin-2-binding protein).  
GN HUP OR HLP OR LBP21 OR RV2986C OR MT3064 OR MTCY349.01.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby J., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 71-86, AND DNA-BINDING.  
RC STRAIN=H37Rv;  
RA Prasad H.K., Annapurua P.S., Day A.B., Tyagi J.S., Jain N.K.,  
RA Savita P.;  
RL Submitted (DEC-1997) to the SWISS-PROT data bank.  
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Z83018; CAB05427.1; -.  
DR EMBL; AE007127; AAK47393.1; -.  
DR HSSP; P02346; 1HUU.  
DR TIGR; MT3064; -.  
DR TuberculList; RV2986c; -.  
DR InterPro; IPR001119; Bac\_DNABind.  
DR InterPro; IPR001386; Histone\_H1/H5.  
DR Pfam; PF00216; Bac DNA binding; 1.  
DR PRINTS; PR00624; HISTONEHS.  
DR ProDom; PD000945; Bac\_DNABind; 1.  
DR SMART; SM00411; BHL; 1.  
DR PROSITE; PS00045; HISTONE LIKE; 1.  
KW DNA-binding; DNA condensation; Repeat; Complete proteome.  
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
FT DOMAIN 101 214 DEGENERATE REPEATS REGION.  
SQ SEQUENCE 214 AA; 22187 MW; CB09AF20FB353544 CRC64;  
-----  
Query Match 33.9%; Score 106; DB 1; Length 214;  
Best Local Similarity 42.4%; Pred. No. 0.03;  
Matches 39; Conservative 5; Mismatches 26; Indels 26; Gaps 4;  
QY 1 AKKYAKK---EKAYAKAKKAEAKA-----AKKA---KAEKKYAKAAKAEKKEAYAA 45  
Db 111 AKKVAKKAPAKATKAATKAATKAPAKAAATKAPAKAAATKAPAKAAATKATKSPAKKVTK 170  
QY 46 A-----EAKYKAEAAKAAKAAVEA 66  
Db 171 AVKKTAVKASVKAATKAPAKAAAKRPAKTA 202  
-----  
Search completed: March 10, 2003, 12:17:09  
Job time : 15.038 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 34.9494 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-4  
Perfect score: 313  
Sequence: 1 AKYAKKEKAYAKAKAEK.....EAKYAKAEAKAAKAAEAYEA 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21.\*

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.5	42.3	394	16	Q8X965
2	126.5	40.4	389	16	Q9CM70
3	124.5	39.8	407	16	Q8ZQT6
4	122.5	39.1	372	2	Q9WXX1
5	122.5	39.1	376	16	Q8ZBC1
6	120	38.3	395	2	Q937K4
7	117	37.4	388	16	Q8ZG22
8	113.5	36.3	275	5	O01395
9	113	36.1	284	10	Q6S794
10	112	35.8	243	16	Q92A67
11	111.5	35.6	239	16	Q8Y5W4
12	110.5	35.3	232	10	Q39576
13	110	35.1	275	10	Q9XHL9
14	110	35.1	1341	16	Q98KG7
15	108.5	34.7	198	4	Q9H8H4
16	108.5	34.7	467	4	Q9H9F1

17	107.5	34.3	227	10	Q9SWU1
18	107.5	34.3	238	10	Q9XHL8
19	107	34.2	200	16	Q8XVW7
20	107	34.2	1701	5	O61164
21	106.5	34.0	236	10	Q9SWU3
22	106.5	34.0	237	10	Q9SWU2
23	106.5	34.0	345	16	Q8Y1F6
24	106.5	34.0	755	10	Q9FP71
25	105	33.5	213	3	Q9UV33
26	104	33.2	298	2	Q52088
27	104	33.2	311	12	Q84528
28	104	33.2	1866	5	Q8T5C8
29	103	32.9	182	2	Q45370
30	102.5	32.7	436	11	Q99KC2
31	102	32.6	224	13	Q90ZD7
32	102	32.6	312	5	Q9UAN1
33	102	32.6	485	10	Q8RXD0
34	102	32.6	924	10	Q9SU08
35	101.5	32.4	629	2	Q9KJ98
36	101.5	32.4	969	5	Q9ND19
37	101.5	32.4	1128	5	Q26947
38	100.5	32.1	166	5	Q99281
39	99.5	31.8	212	3	Q93946
40	99.5	31.8	229	13	Q91369
41	99	31.6	581	16	Q9RU45
42	98.5	31.5	233	5	O18319
43	98.5	31.5	243	5	Q23784
44	98.5	31.5	607	5	Q9V6S7
45	98	31.3	180	5	Q25636

## ALIGNMENTS

## RESULT 1

Q8X965 ID Q8X965 PRELIMINARY; PRT; 394 AA.

AC Q8X965;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Membrane spanning protein, required for outer membrane integrity

DE (Membrane spanning protein TolA).

GN TOLA OR Z0907 OR ECS0774.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OC Escherichia

OX NCBI\_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G., Evans P.S., Gregor J., Kirkpatrick H.A., Rose D.J., Mayhew G.F., Boutin A., Shao Y., Miller L., Posfai G., Hackett J., Klink S., Lim A., Dimalanta E.T., Potamouis K., Grotbeck E.J., Davis N.W., Lin J., Yen G., Schwartz D.C., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature 409:529-533(2001).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12." DNA Res. 8:11-22(2001).

RL DNA Res. 8:11-22(2001).

DR EMBL; AF005252; AAG55075.1; -

DR EMBL; AF002553; BAB34197.1; -





DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE Axoneme-associated protein MGT101(3).  
 GN MGT101(3) OR DHMST101.  
 OS Drosophila hydei (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Neesen J., Heinlein U.A.O., Buenemann H.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).  
 CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
 DR EMBL; U85627; AAB51369.1; -.  
 DR FlyBase; FBgn020732; Dhyd\mst101(3).  
 KW Sperm; Repeat; Multigene family.  
 FT DOMAIN 64 255  
 FT 13 X 16 AA APPROXIMATE TANDEM REPEATS OF  
 FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-  
 FT [AE]-X.  
 SQ SEQUENCE 275 AA; 30436 MW; 76BAA7B2A2DF732C CRC64;  
 Query Match 36.3%; Score 113.5; DB 5; Length 275;  
 Best Local Similarity 52.7%; Pred. No. 0.015;  
 Matches 39; Conservative 3; Mismatches 15; Indels 17; Gaps 4;  
 QY 5 AKKEKAYAKAKAKAKAKAK--AEARKYAKAKAE-----KKEYAAAEAK 49  
 DB 71 AKKEKEAAEKKC-AEAAKKEAAEKKCAEAAKKEQAAQKKCAELAKKEAAEK 129  
 QY 50 YKAEAAKAAKEAA 63  
 DB 130 KCAEAAK-KEKEAA 142  
 RESULT 9  
 ID 065794 PRELIMINARY; PRT; 284 AA.  
 AC 065794;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Histone H1.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98249625; PubMed=9588026;  
 RA Taoka K., Ohtsubo N., Fujimoto Y., Mikami K., Meshi T., Iwabuchi M.;  
 RT "The modular structure and function of the wheat H1 promoter with S  
 RT phase-specific activity";  
 RL Plant Cell Physiol. 39:294-306(1998).  
 DR EMBL; D87064; BAA25203.1; -.  
 DR HSP; P02259; IHST.  
 DR InterPro; IPR000637; AT hook.  
 DR InterPro; IPR001386; Histone H1/H5.  
 DR InterPro; IPR003216; LinkerHist\_N.  
 DR Pfam; PF02178; AT hook; 1.  
 DR Pfam; PF00538; linker histone; 1.  
 DR ProDom; PD000373; LinkerHist\_N; 1.  
 DR SMART; SM00384; AT hook; 1.  
 DR SMART; SM00526; H15; 1.  
 SQ SEQUENCE 284 AA; 29159 MW; 5165A2B922BCBA4E CRC64;  
 Query Match 36.1%; Score 113; DB 10; Length 284;  
 Best Local Similarity 52.5%; Pred. No. 0.017;

Matches 32; Conservative 6; Mismatches 21; Indels 2; Gaps 1;  
 QY 2 KKYAKKEKAYAKAKAEAAK--AKAAKAYAKAAKAEKKEYAAAEAKYKAEAAKAA 59  
 DB 152 KPPAAKKAPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 211  
 QY 60 K 60  
 DB 212 K 212  
 RESULT 10  
 ID 092A67 PRELIMINARY; PRT; 243 AA.  
 AC 092A67;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein lin2055.  
 GN LIN2055.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunsch F., Kurapkai G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.;  
 RT "Comparative genomics of Listeria species";  
 RL Science 294:849-852(2001).  
 DR EMBL; AL596170; CAC97285.1; -.  
 DR ListList; LIN02055; -.  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 243 AA; 25963 MW; 6B2493D143B159D1 CRC64;  
 Query Match 35.8%; Score 112; DB 16; Length 243;  
 Best Local Similarity 47.1%; Pred. No. 0.018;  
 Matches 32; Conservative 11; Mismatches 23; Indels 2; Gaps 1;  
 QY 1 AKKYAKKEKAYAKAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAA 58  
 DB 110 ABEKAAAEKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAA 169  
 QY 59 AKEAAVEA 66  
 DB 170 KEKAAEA 177  
 RESULT 11  
 ID 08Y5W4 PRELIMINARY; PRT; 239 AA.  
 AC 08Y5W4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein lmo1941.  
 GN LMO1941.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.

```

Db      146 KCGEKKKA-AKPAKEKKPKAAKPAKKTPTTKAAAKPKAEKKPKAAAKPKAEKKPKAAK 204
QY      46 AEAKYKAEAAKAAKAAKEAA 63
Db      205 PKAEKKAAKPAKAKSA 222

RESULT 13
Q9XHL9
ID      Q9XHL9      PRELIMINARY;      PRT;      275 AA.
AC      Q9XHL9;
DT      01-NOV-1999 (TrEMBLrel. 12, Created)
DT      01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT      01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE      Histone H1 WH1B.1.
OS      Triticum aestivum (Wheat).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC      Triticeae; Triticum.
NCBI_TaxID=4565;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Steinhoff S.;
RT      "Structural and functional characterization of histone H1 from animal
RT      and plant cells.";
RL      Thesis (1998) , University of Goettingen, Humboldtallee.
DR      EMBL; AF107024; AAD41007.1; -.
DR      HSSP; P02259; 1HST.
DR      InterPro; IPR001386; Histone H1/H5.
DR      InterPro; IPR003216; Linkerhist N.
DR      Pfam; PF00538; linker histone; 1.
DR      PRINTS; PR00624; HISTONEH5.
DR      ProDom; PD000373; Linkerhist N; 1.
DR      SMART; SM00526; H15; 1.
SQ      SEQUENCE      275 AA; 27793 MW; B96DB4B710FF188F CRC64;

Query Match      35.1%; Score 110; DB 10; Length 275;
Best Local Similarity 48.7%; Pred. NO. 0.03;
Matches 37; Conservative 5; Mismatches 20; Indels 14; Gaps

QY      1 AKKYAKKEKAYAKAKK-----AEAKAAKAAKAEAKKYAKA---AKAEKKYAAAE 47
Db      151 AKKTATTTKAKAPAKKSAKPKAKAPAKTKAAAKPKAAKPKAKAPAKTKAAAKPKAAAK 210
QY      48 AK-YKAEAAKAAKAAKEA 62
Db      211 PKGPPKAAKTSAKDA 226

RESULT 14
Q98KG7
ID      Q98KG7      PRELIMINARY;      PRT;      1341 AA.
AC      Q98KG7;
DT      01-OCT-2001 (TrEMBLrel. 18, Created)
DT      01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT      01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE      Hypothetical protein mll1482.
GN      MLL1482.
OS      Rhizobium loti (Mesorhizobium loti).
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Phyllobacteriaceae; Mesorhizobium.
NCBI_TaxID=381;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=MAFF303099;
RX      MEDLINE=21082930; PubMed=11214968;
RX      Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA      Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA      Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA      Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA      Takeuchi C., Yamada M., Tabata S.;
RT      "Complete genome structure of the nitrogen-fixing symbiotic bacterium

```

```
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002997; BAB48847.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1341 AA; 138670 MW; 79C5D8149C969BC8 CRC64;

Query Match 35.1%; Score 110; DB 16; Length 1341;
Best Local Similarity 53.5%; Pred. No. 0.14;
Matches 38; Conservative 9; Mismatches 16; Indels 8; Gaps 4;

Qy 1 AKKYAKKE---KAYAKAK-KAEAKAAKAKAEAKKYAKA---AKAEKKEYAAAEAKYKAE 53
Db 1195 AKAKAEAEQAQAQAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAD 1254
Qy 54 A-AKAAAKEAA 63
Db 1255 ADAKAAAEQQA 1265

RESULT 15
Q9H8H4 PRELIMINARY; PRT; 198 AA.
AC Q9H8H4
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDNA FLJ13629 fis, clone PLACE1011056, weakly similar to histone H1,
DS gonadal.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Iehii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK023691; BAB14642.1; -.
DR InterPro; IPR000637; AT hook.
DR SMART; SM00384; AT hook; 1.
SQ SEQUENCE 198 AA; 21353 MW; 46F43602BC12487B CRC64;

Query Match 34.7%; Score 108.5; DB 4; Length 198;
Best Local Similarity 50.7%; Pred. No. 0.029;
Matches 36; Conservative 9; Mismatches 17; Indels 9; Gaps 4;

Qy 1 AKKYAKKEKAYAKAKAEAKAK---KAKA---EAKKYAKAA--KAEKKEYAR-AEAKYK 51
Db 17 ARTLAKARQAQAQVARTQAQAARAKAKAKAALVKAQAKAKAAQVKAQKVMAAFAKAKAK 76
Qy 52 AEAATAAAKEA 62
Db 77 AKAVRAKAKVA 87
```

Search completed: March 10, 2003, 12:25:56  
Job time : 35.9494 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 34.3924 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-4  
Perfect score: 313  
Sequence: 1 AKYAKKAYAKAKAEAK.....EAKYKAEAKAAKAEAYEA 66

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_101002.\*

- 1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	313	100.0	66	21	AA82574 Copolymer molecula
2	282.5	90.3	77	21	AA82575 Copolymer molecula
3	268	85.6	86	21	AA82576 Copolymer molecula
4	228.5	73.0	109	21	AA82577 Copolymer molecula
5	198	63.3	56	21	AA82573 Copolymer molecula
6	129.5	41.4	154	11	AA806445 Recombinant copoly
7	128	40.9	100	21	AA894499 Peptide #10 used i
8	128	40.9	100	21	AA895044 Amino acid polymer
9	128	40.9	100	22	AAU04289 Poly-Lys-Ala used
10	128	40.9	100	22	AA845852 Nucleic acid trans

11	122	39.0	106	11	AA806446 Recombinant copoly
12	120.5	38.5	45	21	AA82572 Copolymer molecula
13	111.5	35.6	239	23	AB849123 Listeria monocytog
14	110.5	35.3	205	21	AB20575 Mycobacterium bovi
15	108.5	34.7	198	22	AB95499 Human protein sequ
16	108.5	34.7	279	22	AAU03592 Human DNA modifica
17	108.5	34.7	467	22	AB94309 Human protein sequ
18	106	33.9	214	20	AA82571 M. tuberculosis hi
19	106	33.9	214	21	AA82571 M. tuberculosis hi
20	105.5	33.7	35	21	AA82571 Copolymer molecula
21	105	33.5	1507	21	AB24128 plasmidium Yoelii
22	104.5	33.4	223	20	AA82571 Amino acid sequenc
23	104	33.2	140	13	AA82571 Synthetic helical
24	103	32.9	334	22	AB828693 Novel human diagno
25	102.5	32.7	433	18	AA82571 Saccharom
26	102.5	32.7	433	20	AA82571 Human zuotin prote
27	102.5	32.7	433	22	AA82571 S cerevisiae apopt
28	102	32.6	46	18	AA82571 Nucleic acid (NA)
29	102	32.6	299	22	AB863276 Drosophila melanog
30	101.5	32.4	643	16	AA84568 Trypanosoma cruzi
31	101	32.3	829	22	AA82571 Haemophilus influ
32	100.5	32.1	564	16	AA84565 Trypanosoma cruzi
33	99	31.6	859	23	AA82571 Bacillus sp alkali
34	98.5	31.5	607	22	AB862173 Drosophila melanog
35	98	31.3	219	21	AA82571 Quail H1 histone p
36	95.5	30.5	472	16	AA84569 Trypanosoma cruzi
37	95	30.4	160	18	AA82571 Streptococcus pneu
38	95	30.4	427	22	AA82571 C albicans apoptos
39	95	30.4	617	22	AB82571 Peptide #2951 enco
40	95	30.4	617	22	AB82571 Protein #2883 enco
41	95	30.4	617	22	AA82571 Human brain expres
42	95	30.4	617	22	AA82571 Peptide #2892 enco
43	95	30.4	617	22	AA82571 Peptide #2869 enco
44	95	30.4	1026	22	AA82571 Human protein SEQ
45	94	30.0	46	13	AA82571 High affinity macr

ALIGNMENTS

RESULT 1

AA82574  
ID AA82574 standard; peptide; 66 AA.

XX AA82574;

XX AC (first entry)

DT 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX WO200018794-A1.

XX 06-APR-2000.

XX 24-SEP-1999; 99WO-US22402.

XX 25-SEP-1998; 98US-0101693.

XX (YEDA ) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.





Db 61 YKAERKAYAKAEKKEKYEAAAEAKKAEAAKAYKAEAAKAEAAKAEAAEAA 109

## RESULT 5

AY82573

ID AY82573 standard; peptide; 56 AA.

XX AC AAY82573;

XX DT 28-JUL-2000 (first entry)

XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.

XX OS Unidentified.

XX PN WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.  
XX PA (TEVA-) TEVA PHARM USA INC.

XX PI Gad A, Lis D;

XX DR WPI; 2000-317499/27.

XX PT Copolymer 1 related polypeptides used as molecular weight markers for  
XX PT glatiramer acetate and for treatment and prevention of immune diseases

XX PS Claim 10; Page 14; 72pp; English.

XX CC AY82571 to AY82577 represent specifically claimed copolymer molecular  
XX CC weight TV-marker polypeptides from the present invention. The present  
XX CC invention describes polypeptides (I) for determining the molecular  
XX CC weight of a copolymer (CP), which has an identified molecular weight  
XX CC and an amino acid composition corresponding to the copolymer. The  
XX CC polypeptides of the invention are used as molecular weight markers for  
XX CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
XX CC used for treating and preventing immune diseases in a mammal. Autoimmune  
XX CC diseases which may be treated include either cell-mediated or  
XX CC antibody-mediated diseases. Such diseases include arthritic conditions,  
XX CC demyelinating diseases and inflammatory conditions, e.g. multiple  
XX CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
XX CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
XX CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
XX CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
XX CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
XX CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
XX CC lupus erythematosus. Mediated-mediated diseases which can be treated  
XX CC include host-versus-graft disease, graft-versus-host disease, and  
XX CC delayed-type hypersensitivity. The polypeptides of the invention have  
XX CC defined molecular weights and physical properties which are analogous to  
XX CC glatiramer acetate molecules, which makes them ideal for use as  
XX CC molecular weight markers.

XX SQ Sequence 56 AA;

Query Match 63.3%; Score 198; DB 21; Length 56;

Best Local Similarity 80.3%; Pred. No. 5.9e-13;  
Matches 53; Conservative 2; Mismatches 1; Indels 10; Gaps 5;

OY 1 AKKYAKKKKAYAKAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAK 60  
|||||

Db 1 AKKYAKKKKAY--AKKAE--KAAK--KAEAKAY--KAEAKKK--AEAKYKAEAAKAAK 50  
|||||

OY 61 EAAVEA 66  
|||||

Db 51 EAAVEA 56  
|||||

## RESULT 6

AAR06445

ID AAR06445 standard; protein; 154 AA.

XX AC AAR06445;

XX DT 03-JAN-1991 (first entry)

XX DE Recombinant copolymer 1-77, myelin basic protein analogue.

XX KW Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
KW immunological activity; autoimmune encephalomyelitis;  
KW multiple sclerosis;

XX OS Synthetic.

XX PN EP383620-A.

XX PD 22-AUG-1990.

XX PF 16-FEB-1990; 90EP-0301700.

XX PR 07-FEB-1990; 90US-0473845.

XX PR 17-FEB-1989; 89US-0312541.

XX PA (REPL-) REPLIGEN CORP.

XX PI Cook KS;

XX DR WPI; 1990-255848/34.

XX DR N-PSDB; AAQ05664.

XX PT Producing genes encoding random polymers of aminoacid(s) - for  
XX PT producing recombinant polypeptide(s) with biological and/or  
XX PT immunological activity

XX Dislosure; Fig 11; 25pp; English.

XX CC To improve the expression of rCOP-1 polypeptides in E. coli, genes  
XX CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pEG3-2deltaN  
XX CC (deposited: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
XX CC express Protein A. The resulting plasmids encode fusion proteins  
XX CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
XX CC A methionine residue occurs between the Protein A and rCOP-1  
XX CC sequences, originating from the 5' linker sequence, in order that  
XX CC the COP-1 polypeptide may be cleaved from the fusion protein.  
XX CC rCOP-1-77 contains oligonucleotide duplexes encoding the following  
XX CC segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue  
XX CC is left behind following CNBr cleavage of the fusion protein.  
XX CC The product prevents or arrests experimental autoimmune  
XX CC encephalomyelitis. They are used to prevent, arrest or control a  
XX CC demyelinating disorder, e.g. multiple sclerosis. They may also  
XX CC be used as additives to hair care products to confer beneficial  
XX CC effects on damaged hair or as supplements for diets deficient in  
XX CC certain amino acids.  
XX CC See also AAQ05665.

XX SQ Sequence 154 AA;

Query Match 41.4%; Score 129.5; DB 11; Length 154;  
Best Local Similarity 48.8%; Pred. No. 1e-05;



```
Best Local Similarity 62.7%; Pred. No. 9.3e-06;
Matches 42; Conservative 5; Mismatches 16; Indels 4; Gaps 4;

Qy 1 AKKYAK-KEKAYAK-KAEAKAKAKAKAEAKKYAKA-AKAEKKEYAAAEAKYKAE-AK 56
Db 2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61

Qy 57 AAKEAA 63
Db 62 AKAKAKA 68

RESULT 10
AAB45852
ID AAB45852 standard; Peptide; 100 AA.
XX
AC AAB45852;
XX
DT 21-MAR-2001 (first entry)
DE Nucleic acid transporter system peptide ligand SEQ ID NO 64.
XX
KW Nucleic acid delivery; nucleic acid transporter system; hormone; enzyme;
growth factor; clotting factor; apolipoprotein; receptor; drug; oncogene;
tumor antigen; tumor suppressor; viral antigen; parasitic antigen;
bacterial antigen.
XX
OS Unidentified.
XX
PN US6150168-A.
XX
PD 21-NOV-2000.
XX
PF 05-JUN-1995; 95US-0460971.
XX
PR 14-DEC-1993; 93US-0167641.
PR 20-MAR-1992; 92US-0855389.
PR 19-MAR-1993; 93WO-US02725.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Gottchalk S, Sparrow J, Cristiano RJ, Smith LC, Woo SLC;
WPI; 2001-049093/06.
XX
PT Nucleic acid transporter system for delivering nucleic acid into a
cell, useful for delivering proteins and polypeptides to cells,
including growth factors, enzymes, hormones, and tumor suppressors -
XX
PS Disclosure; Column 125-126; 105pp; English.
XX
CC This invention describes a novel system (I) for delivering a nucleic acid
to a cell, comprising a binding complex comprising a ligand binding
molecule noncovalently bound to a nucleic acid and covalently linked to a
surface ligand, and a second binding complex comprising a second binding
molecule noncovalently bound to a nucleic acid and covalently linked to a
nuclear ligand. The complexes are simultaneously bound to the nucleic
acid. The nucleic acid transporter system can also be used in a method
for the in vivo targeting of the insertion of DNA into a cell. It can
also be used in processes for producing transformed cell lines. The
system can be used to deliver a variety of proteins and polypeptides,
such as hormones, growth factors, enzymes, clotting factors,
apolipoproteins, receptors, drugs, oncogenes, tumor antigens, tumor
suppressors, viral antigens, parasitic antigens, and bacterial antigens.
The transporter system uses lysis agents to overcome the problems of
endosomal/lysosomal degradation seen with prior art systems.
XX
SQ Sequence 100 AA;

Query Match 40.9%; Score 128; DB 22; Length 100;
Best Local Similarity 62.7%; Pred. No. 9.3e-06;
Matches 42; Conservative 5; Mismatches 16; Indels 4; Gaps 4;

Qy 1 AKKYAK-KEKAYAK-KAEAKAKAKAKAEAKKYAKA-AKAEKKEYAAAEAKYKAE-AK 56
Db 2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61

Qy 57 AAKEAA 63
Db 62 AKAKAKA 68

RESULT 9
AAU04289
ID AAU04289 standard; Peptide; 100 AA.
XX
AC AAU04289;
XX
DT 23-OCT-2001 (first entry)
DE Poly-Lys-Ala used in nucleic acid transporter system.
XX
KW Nucleic acid transport; cytosol; ligand; lysis agent; spacer molecule;
gene therapy; hepatocyte; muscle; bone forming cell.
XX
OS Synthetic.
XX
FH Location/Qualifiers
FT Misc-difference 3..100
FT /note= "Lys-Ala in positions 3-100 may be present or
FT absent"
XX
US6177554-B1.
XX
PD 23-JAN-2001.
XX
PF 05-JUN-1995; 95US-0462040.
XX
PR 14-DEC-1993; 93US-0167641.
PR 20-MAR-1992; 92US-0855389.
PR 19-MAR-1993; 93WO-US02725.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Woo SLC, Smith LC, Cristiano RJ, Gottchalk S, Sparrow J;
WPI; 2001-365933/38.
XX
PT Nucleic acid transport system, useful for creating transgenic animals
for assessing human disease such as cancer in an animal model -
XX
PS Disclosure; Column 131; 111pp; English.
XX
CC The sequence represents poly-Lys-Ala, used to bind nucleic acid in a
nucleic acid transporter system. The nucleic acid transporter system uses
nucleic acid binding complexes containing surface ligands which are
capable of binding to a cell surface receptor and entering the cell
through cytosol. The compounds of the invention are either ligands,
binding molecules (surface ligands), lysis agents, spacer molecules or
their intermediates. The ligands, binding molecules, lysis agents and
spacer molecules are used in nucleic acid transporter systems to deliver
nucleic acid into specific cells e.g. in gene therapy to deliver nucleic
acid into hepatocytes, muscle cells or bone forming cells.
XX
SQ Sequence 100 AA;

Query Match 40.9%; Score 128; DB 22; Length 100;
Best Local Similarity 62.7%; Pred. No. 9.3e-06;
Matches 42; Conservative 5; Mismatches 16; Indels 4; Gaps 4;

Qy 1 AKKYAK-KEKAYAK-KAEAKAKAKAKAEAKKYAKA-AKAEKKEYAAAEAKYKAE-AK 56
Db 2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61
```











THIS PAGE BLANK (USPTO)







```
QY 61 EAAVEA 66
|||||
Db 30 EAAVEA 35

RESULT 10
US-10-051-643-201
; Sequence 201, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION: System using Mycobacterium Vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-201

Query Match 33.4%; Score 104.5; DB 9; Length 223;
Best Local Similarity 52.2%; Pred. No. 0.0032;
Matches 36; Conservative 4; Mismatches 20; Indels 9; Gaps 3;

QY 1 AKKYAKKEKAYAK---AKKAAKAKKAKAEAKKYAKAAKAEKKEKVEAAAPAKYAAEAAKA 57
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 AKKAAKAAKAAKAPKAPKAPKATKAAAPKATAAK--KAAPAKK---ATAAKKAAAPAKKA 174

QY 58 AAKEAAVEA 66
|||||:|
Db 175 PAKKAATKA 183

RESULT 11
US-09-815-242-11216
; Sequence 11216, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zvakind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
```

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11216
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11216

Query Match 32.3%; Score 101; DB 10; Length 829;
Best Local Similarity 34.8%; Pred. No. 0.026;
Matches 23; Conservative 20; Mismatches 23; Indels 0; Gaps 0;

QY 1 AKKYAKKEKAYAKAEAKAEAKKAAKAAKAEKKEKVEAAAPAKYAAEAAKAAKAAK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 69 AEQKAAEERAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 128

QY 61 EAAVEA 66
:|:|
Db 129 KAEQEA 134

RESULT 12
US-09-820-843A-27
; Sequence 27, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: tola protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|9856364
US-09-820-843A-27

Query Match 31.2%; Score 97.5; DB 9; Length 356;
Best Local Similarity 43.8%; Pred. No. 0.023;
Matches 39; Conservative 8; Mismatches 19; Indels 23; Gaps 5;

QY 1 AKKYAKKEKAYAKAEAKAEAKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 131 AEQKAAEERAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 190

QY 50 YKAE---AKA-----AAKEAAVEA 66
|||||:|
Db 191 AKREKVEAKAEQERLAKEKAAKAEADKA 219

RESULT 13
US-09-864-761-36182
; Sequence 36182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
```



THIS PAGE BLANK (USPTO)









```

RESULT 7
US-09-041-889-27
/ Sequence 27, Application US/09041889
/ Patent No. 6033864
/ GENERAL INFORMATION:
/ APPLICANT: Braun, Jonathan
/ APPLICANT: Crahn, Offer
/ TITLE OF INVENTION: Diagnosis, P
/ TITLE OF INVENTION: Ulcerative UC
/ TITLE OF INVENTION: Microbial UC
/ NUMBER OF SEQUENCES: 41
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Campbell & Flores L
/ STREET: 4370 La Jolla Village
/ CITY: San Diego
/ STATE: California
/ COUNTRY: USA
/ ZIP: 92122
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTED: IBM PC compatible
/

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-27

Query Match 33.9%; Score 106; DB 3; L
Best Local Similarity 42.4%; Pred. NO. 0.0014;
Matches 39; Conservative 5; Mismatches 22;

QY 1 AKYAK--EKAYAKAKAEAKA-----AKKA--KAEAKA
||| ||| :| ||| ||| ||| ||| |||
DB 111 AKKVAKKAPAKKATKAAKKKAATKAPARKAATKAPAKKAATKAPAKKAATKAPAKA

QY 46 A-----EAKYKAEAKAKAEAAEYA 66
:| ||| ||| ||| ||| :|
DB 171 AKKTVAKSVKKAATKAPAKKAARPATKA 202

RESULT 8
US-08-929-329-5
; Sequence 5, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for
; TITLE OF INVENTION: Vaccine Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

```

```
; LENGTH: 1507 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; US-08-929-329-5

Query Match 33.5%; Score 105; DB 3; Length 1507;
Best Local Similarity 50.8%; Pred. No. 0.012;
Matches 33; Conservative 6; Mismatches 14; Indels 12; Gaps 3;

Qy 1 AKYAKKEKAYAK-----AKAAKAKAEAKYAKAAKAEKAYAAAEAKYAAEA 55
Db 1179 AKAAEERKRIEAKAEERKKIEAKAEERK-----AEAVK---AEAKKAEAA 1231
Qy 56 KAAK 60
Db 1232 KKALK 1236

RESULT 9
US-09-855-855-201
; Sequence 201, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Compounds and Methods for
; TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 201:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; US-09-095-855-201

Query Match 33.4%; Score 104.5; DB 4; Length 223;
Best Local Similarity 52.2%; Pred. No. 0.002;
Matches 36; Conservative 4; Mismatches 20; Indels 9; Gaps 3;

Qy 1 AKYAKKEKAYAK---AKKAEAKAAKAKAEAKYAKAAKAEKAYAAAEAKYAAEA 57
Db 121 AKAAAKKAAPAKKAAATKAAPAKKATAAK--KAAPAKK---ATAAKKAAPAKKA 174
Qy 58 AAKEAAVEA 66
Db 175 PAKKAATKA 183

RESULT 10
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; US-09-205-426-201

Query Match 33.4%; Score 104.5; DB 4; Length 223;
Best Local Similarity 52.2%; Pred. No. 0.002;
Matches 36; Conservative 4; Mismatches 20; Indels 9; Gaps 3;

Qy 1 AKYAKKEKAYAK---AKKAEAKAAKAKAEAKYAKAAKAEKAYAAAEAKYAAEA 57
Db 121 AKAAAKKAAPAKKAAATKAAPAKKATAAK--KAAPAKK---ATAAKKAAPAKKA 174
Qy 58 AAKEAAVEA 66
Db 175 PAKKAATKA 183

RESULT 11
US-08-346-849-2
; Sequence 2, Application US/08346849
; Patent No. 5670483
; GENERAL INFORMATION:
; APPLICANT: Zhang, Shuguang
; APPLICANT: Lockshin, Curtis
; APPLICANT: Rich, Alexander
; APPLICANT: Holmes, Todd
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-293-284A-2

Query Match 32.7%; Score 102.5; DB 2; Length 433;
Best Local Similarity 41.0%; Pred. No. 0.0061;
Matches 32; Conservative 9; Mismatches 22; Indels 15; Gaps
349

QY 2 KTVAKEK-----AYAKK-KAEKAKKAKAEAKKYAKAKAEKKEAYAA 46
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 KEKEKKEKKRWERAGARAEAKAEAKAEAKAESEAKANASAKADKKKAKEAAKA 349
; : : : : : : : : : : : : : : : : : : : : : : :

QY 47 EAKYKAEAKAAKAEAY 64
; : : : : : : : : : : : : : : : : : : : : : : :

Db 350 AKKKNKRAIRNSAKEADY 367
; : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-08-216-894-2
; Sequence 2, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Ocau, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Benti, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-2

Query Match 32.4%; Score 101.5; DB 2; Length 564;

```

Best Local Similarity 46.0%; Pred. No. 0.0099;  
Matches 29; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAKAEAKKAKAEAKKYAKAKA-EKKEYAAAEAKKYAKAEAKAAK 60  
Db 331 QRAAEAAKAVETEKQRAAEATKVAAEKRAAEAKAVETEKQRAAEATKVAAEKQKAA 390  
Qy 61 EAA 63  
Db 391 EAA 393

## RESULT 14

US-09-115-746-2  
Sequence 2, Application US/09115746  
Patent No. 6228601  
GENERAL INFORMATION:  
APPLICANT: Kirchhoff, Louis V.  
APPLICANT: Otsu, Keiko  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/115,746  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/216,894  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 564 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-115-746-2

Query Match 32.4%; Score 101.5; DB 4; Length 564;  
Best Local Similarity 46.0%; Pred. No. 0.0099;  
Matches 29; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAKAEAKKAKAEAKKYAKAKA-EKKEYAAAEAKKYAKAEAKAAK 60  
Db 331 QRAAEAAKAVETEKQRAAEATKVAAEKRAAEAKAVETEKQRAAEATKVAAEKQKAA 390  
Qy 61 EAA 63  
Db 391 EAA 393

## RESULT 15

US-08-216-894-8  
Sequence 8, Application US/08216894  
Patent No. 5876734  
GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.  
APPLICANT: Otsu, Keiko  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,894  
FILING DATE: 24-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-216-894-8

Query Match 32.4%; Score 101.5; DB 2; Length 643;  
Best Local Similarity 46.0%; Pred. No. 0.011;  
Matches 29; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAKAEAKKAKAEAKKYAKAKA-EKKEYAAAEAKKYAKAEAKAAK 60  
Db 331 QRAAEAAKAVETEKQRAAEATKVAAEKRAAEAKAVETEKQRAAEATKVAAEKQKAA 390  
Qy 61 EAA 63  
Db 391 EAA 393

Search completed: March 10, 2003, 12:30:03  
Job time : 12.6962 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 22.5802 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989A-5  
Perfect score: 366  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYKAEAAKAAKEAAEYEA 77

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150.5	41.1	421	2 JV0057	tola protein - Esc
2	148	40.4	394	2 F90725	membrane spanning
3	148	40.4	394	2 G85576	membrane spanning
4	146	39.9	376	2 AG0592	tola protein [impo
5	141.5	38.7	347	2 E83525	Tola protein PA097
6	141	38.5	372	2 G64064	outer membrane int
7	138	37.7	1701	2 T09127	probable erythrocy
8	136	37.2	210	2 A25550	histone H1 - sea u
9	134.5	36.7	219	2 E60110	repetitive protein
10	134.5	36.7	231	2 S9589	histone H1 - Chlam
11	134	36.6	1128	2 T30296	R27-2 protein - Tr
12	133	36.3	328	2 A44993	cytosolic repetiti
13	130.5	35.7	388	2 AC0138	TOLA colicin impor
14	130	35.5	1390	2 S51364	sperm tail-specifi
15	129.5	35.4	214	2 G70673	probable hupB - My
16	128.5	35.1	211	2 A8100	histone H1-beta, e
17	127.5	34.8	344	2 S34153	met101-1 protein -
18	127	34.7	217	2 A26721	histone H1-gamma,
19	123	33.6	248	1 H5UR1P	histone H1, gonada
20	123	33.6	433	2 S25194	zucotin - yeast (Sa
21	122.5	33.5	206	2 S9388	histone H1 - sea u
22	122.5	33.5	220	2 A28456	histone H1.10 - ch
23	122.5	33.5	311	2 T17698	hypothetical prote
24	122	33.3	243	2 A81689	hypothetical prote
25	121.5	33.2	239	2 A81317	hypothetical prote
26	121.5	33.2	384	2 B43592	outer membrane pro
27	120.5	32.9	284	2 T06241	histone H1 (clone
28	118.5	32.4	241	2 JN0748	histone H1-II - Vo
29	118	32.2	206	1 HSTR1R	histone H1 - rainb

30	118	32.2	1403	2 T11583	probable translati
31	117.5	32.1	265	2 S19113	cgr-4 protein - C
32	117.5	32.1	288	2 T06257	histone H1 (clone
33	115.5	31.6	225	2 B28456	histone H1.11L - c
34	115.5	31.6	924	2 T06636	hypothetical prote
35	114	31.1	581	2 E75383	conserved hypother
36	113.5	31.0	224	2 D28456	histone H1.03 - ch
37	112.5	30.7	218	1 HSC11	histone H1.02 - ch
38	112	30.6	218	2 S01262	histone H1 - musco
39	111.5	30.5	325	2 F71283	outer membrane ant
40	111.5	30.5	356	2 A82152	tola protein VC183
41	111	30.3	182	2 S61926	histone H1 homolog
42	111	30.3	208	2 T23778	histone H1.1 - Cae
43	111	30.3	218	2 A23055	histone H1.01 - ch
44	110.5	30.2	226	1 S51660	histone H1-5 [vali
45	110.5	30.2	771	1 A33430	h-caldesmon - chic

ALIGNMENTS

RESULT 1

JV0057  
tola protein - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002  
C:Accession: JV0057; B64810  
R:Levengood, S.K.; Webster, R.E.  
J: Bacteriol. 171, 6600-6609, 1989  
A:Title: Nucleotide sequences of the tola and tolB genes and localization of their products  
A:Reference number: JV0057; MUID:90078104; PMID:2687247  
A:Accession: JV0057  
A:Molecule type: DNA  
A:Residues: 1-421 <LEV>  
A:Cross-references: GB:M28232; NID:G148018; PIDN:AAA24683.1; PID:G148019  
A:Experimental source: strain JM105  
A:Note: The authors translated the initiation codon GTG for residue 1 as Val  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: B64810  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-421 <BLAT>  
A:Cross-references: GB:AE000177; GB:U00096; NID:G1786955; PIDN:AAC73833.1; PID:G1786950;  
A:Experimental source: strain K-12, substrain MG1655  
C:Comment: tola and tolB proteins are necessary for colicins E2, E3, A, and K to reach t  
C:Genetics:  
A:Gene: tola  
A:Map position: 17 min  
A:Start codon: GTG  
C:Keywords: nucleotide binding; P-loop; transmembrane protein  
F:14-34/Domain: transmembrane #status predicted <MSS>  
F:78-301/Domain: helical #status predicted <HSR>  
P:355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 41.1%; Score 150.5; DB 2; Length 421;  
Best Local Similarity 53.7%; Pred. No. 0.00019;  
Matches 44; Conservative 8; Mismatches 23; Indels 7; Gaps 2;

Oy	3	KYAKKEKAYAKAEKAAKAYKAAEKKAEEAKKAAEKYAKVAK-----AKAEKK-EYA 55
Db	130	KQKQEEAAKAAADAKAKAEADAKAEAAKAAADAKKAAEAAKAAAEAKAAEAKAAEA 189
Oy	56	AAEAKYKAEAAKAAAEKAAEYEA 77
Db	190	AAALKKAAEAAEAAAEAKKA 211

RESULT 2

F90725



R; Knowles, J.A.; Childs, G.J.  
Nucleic Acids Res. 14, 8121-8133, 1986  
A>Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus and

A;Reference number: A25550; MUID:87040778; PMID:3022245  
A;Accession: A25550  
A:Molecule type: DNA  
A;Residues: 1-210 <KNO>  
A;Cross-references: GB:X04498; NID:g9616; PIDN:CAA28177.1; PID:g9617  
C;Superfamily: histone H1  
C;Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 37.2%; Score 136; DB 2; Length 210;  
Best Local Similarity 53.3%; Pred.No. 0.0013;  
Matches 40; Conservative 7; Mismatches 26; Indels 2; Gaps 2;

Qy 2 KYAKKEKAYAKKAQAQKAAEAKYKAEAK-KKAQAQAQKYVA-QAAEAQEKEYAAAAEA 59  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 96 KTEAQAPAAAQYKALAAKKKEQEKKAATKARKEKLAAKQAQKAAKVVPAAKAKK 155  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 60 KYKAPAAKAAAKAEAA 74  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 156 PAKTAARKPAAKKAA 170  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 9  
E60110  
repetitive protein antigen 27 - Trypanosoma cruzi (fragments)  
C;Species: Trypanosoma cruzi  
C;Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 02-Jun-2000  
C;Accession: E60110  
R;Hofft, D.F.; Kim, K.S.; Otsu, K.; Moser, D.R.; Yost, W.J.; Blumin, J.H.; Donelson, J.E.,  
Infect. Immun. 57, 1959-1967, 1989  
A>Title: Trypanosoma cruzi expresses diverse repetitive protein antigens.  
A;Reference number: A60110; MUID:89277508; PMID:2659529  
A;Accession: E60110  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A;Residues: 1-219 <HOF>  
C;Comment: This protein contains a series of tandem repeats, each fourteen residues in l  
C;Superfamily: neurofilament triplet H protein  
C;Keywords: tandem repeat

Query Match 36.7%; Score 134.5; DB 2; Length 219;  
Best Local Similarity 50.7%; Pred.No. 0.0018;  
Matches 38; Conservative 8; Mismatches 28; Indels 1; Gaps 1;

Qy 1 AKKYAKKEKAYAKKAQAQKAAEAKYKAEAKKKAQAQAQKYVAKAAK-AEQKEYAAAAEA 59  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 70 ATKVAGDEKQAAREATKVAKVAEAEKQAREATKVAAEAEKQAAREATKVAAEAEKQAAREAA 129  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 60 KYKAAQAQAQAQKAA 74  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 130 TKVAEAGKQKAAEEA 144  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 10  
S59589  
histone H1 - Chlamydomonas reinhardtii  
C;Species: Chlamydomonas reinhardtii  
C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 23-Jul-1999  
C;Accession: S59589; S62122  
R;Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.  
Curr. Genet. 28, 333-345, 1995  
A>Title: The organization structure and regulatory elements of Chlamydomonas histone gen  
A;Reference number: S59581; MUID:96120862; PMID:8590479  
A;Accession: S59589  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A;Residues: 1-231 <FAB>  
A;Cross-references: EMBL:U16726  
A>Note: the authors did not translate the codon for residue 1  
R;Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.  
submitted to the EMBL Data Library, October 1994

Query Match 35.5%; Score 130; DB 2; Length 1390;  
Best Local Similarity 51.7%; Pred. No. 0.015;



THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 17.5443 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-5

Perfect score: 366

Sequence: 1 AKYAKKEKAYAKKAERAAK.....EAKYAKBAKAAKAAKAAAYEA 77

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150.5	41.1	421	1 TOLA_ECOLI	P19934 escherichia
2	141.5	38.7	347	1 TOLA_PSEAE	P50600 pseudomonas
3	141	38.5	372	1 TOLA_HAEIN	P44678 haemophilus
4	136	37.2	210	1 H1_LYTP1	P06144 lytechinus
5	130	35.5	205	1 DBH_MYCBO	Q9xb18 mycobacteri
6	130	35.5	1391	1 MST2_DROHY	Q08696 drosophila
7	129.5	35.4	214	1 DBH_WCTU	P95109 mycobacteri
8	129	35.2	208	1 DBH_MTCSM	Q9zhcs mycobacteri
9	127.5	34.8	344	1 MST1_DROHY	Q08695 drosophila
10	127	34.7	217	1 HIG_STRPU	P07796 strongyloce
11	125.5	34.3	211	1 H1B_STRPU	P15869 strongyloce
12	125.5	34.3	349	1 RS6_AEDAL	Q9u762 aedes albop
13	123	33.6	248	1 H1_PARAN	P02256 paracitrus
14	123	33.6	346	1 RS6_AEDAE	Q9u761 aedes aegyp
15	123	33.6	433	1 ZUO1_YEAST	P32527 saccharomyc
16	122.5	33.5	219	1 H101_CHICK	P08286 gallus gall
17	121.5	33.2	384	1 TWBP_TREPH	P29720 treponema p
18	120.5	32.9	221	1 H1C_CHITE	P40277 chironomus
19	118.5	32.4	240	1 H12_VOLCA	Q08865 volvox cart
20	118	32.2	206	1 H1_ONCMY	P06350 oncorhynch
21	118	32.2	237	1 H1E_CHITE	P40278 chironomus
22	118	32.2	1403	1 YDF3_SCHPO	Q10475 schizosacch
23	115.5	31.6	224	1 H1L1_CHICK	P08287 gallus gall
24	115	31.4	232	1 H1B_CHITE	P40276 chironomus
25	114	31.1	233	1 H1L1_GLYSA	P40266 glyptotendi
26	113.5	31.0	223	1 H103_CHICK	P08285 gallus gall
27	113.5	31.0	917	1 IF2_PROVU	Q9zf22 proteus vul
28	113	30.9	238	1 H1_WHEAT	P27806 triticum ae
29	112.5	30.7	217	1 H1_CHICK	P09987 gallus gall
30	112	30.6	217	1 H1_ANAPL	P09426 anas platyr
31	111.5	30.5	139	1 ASR_KLEPN	Q93mh5 klebsiella
32	111.5	30.5	325	1 TWBP_TREPA	P19649 treponema p
33	111	30.3	207	1 H1L1_CAEEL	P10771 caenorhabdi

RESULT 1				
TOLA_ECOLI				
ID	TOLA_ECOLI	STANDARD;	PRT;	421 AA.
AC	P19934;			
DT	01-FEB-1991	(Rel. 17, Created)		
DT	01-FEB-1991	(Rel. 17, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	TOLA protein.			
GN	TOLA OR CIM OR EXCC OR LKY OR B0739.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision;			
OC	Escherichia.			
OX	NCBI_Taxid=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JM105;			
EX	MEDLINE=90078104; PubMed=2687247;			
RA	Levensood S.K., Webster R.E.;			
RT	"Nucleotide sequences of the tola and tolB genes and their products, components of a multistep translocation system in Escherichia coli.";			
RL	J. Bacteriol. 171:6600-6609(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
EX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Fiske S.H., Raskin D., Riley M., Collado-Vides J., Glasner J.D., Rood J., Mayhew G.P., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
EX	MEDLINE=97061202; PubMed=8905232;			
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yanamoto Y., Yano M., Horiuchi T.;			
RT	"A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map."			
RL	DNA Res. 3:137-155(1996).			
RN	[4]			
RP	DOMAINS.			
EX	MEDLINE=91296736; PubMed=2068069;			
RA	Levensood S.K., Beyer W.F. Jr., Webster R.E.;			
RT	"Tola: a membrane protein involved in colicin uptake contains an extended helical region.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).			
RN	[5]			
RP	INTERACTION WITH PORINS.			
EX	MEDLINE=97133271; PubMed=8978668;			
RA	Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,			

34	111	30.3	218	1	H101_CHICK	P08284 gallus gall
35	110.5	30.2	225	1	H15_HUMAN	P16401 homo sapien
36	110.5	30.2	771	1	CALD_CHICK	P12957 gallus gall
37	110	30.1	233	1	H11_GLYBA	P40263 glyptotendi
38	109.5	29.9	194	1	H11_SALTR	P02254 salmo trutt
39	109.5	29.9	209	1	H1A_XENLA	P06892 xenopus lae
40	109	29.8	232	1	H1A_CHITE	P40275 chironomus
41	108	29.5	833	1	IF2_PASMU	P57873 pasteurella
42	107.5	29.4	218	1	H11R_CHICK	P08288 gallus gall
43	106.5	29.1	90	1	H162_TRYCR	P40274 trypanosoma
44	105.5	28.8	235	1	H1E_CHIPA	P40262 chironomus
45	105	28.7	219	1	H1B_XENLA	P06893 xenopus lae

## ALIGNMENTS

```

RX MEDLINE=97113525; PubMed=8955385;
RT Dennis J.J., Lafontaine E.R., Sokol P.A.;
RI "Identification and characterization of the tolQRA genes of
RL Pseudomonas aeruginosa.";
RN J. Bacteriol. 178:7059-7068(1996).
RP [2]
RR REVISIONS TO N-TERMINUS.
RS Duan K., Sokol P.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RC Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brady L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RL opportunistic pathogen."
RN Nature 406:959-964(2000).
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39558; AAC44660.2; -.
DR EMBL; AE004530; AGG04360.1; -.
RW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).
FT DOMAIN 209 216 POLY-ALA.
SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 38.7%; Score 141.5; DB 1; Length 347;
Best Local Similarity 44.1%; Pred. NO. 0.00043;
Matches 45; Conservative 9; Mismatches 19; Indels 29; Gaps 3;

QY 1 AKKYAKKRAYAKKAKKAKEAKA-----YKAAE-A 32
   | : : ||||| : : |||||
Db 120 ARKAEQAQKAEAKKADENKAEKAEKAEQQKQNDIAKKRAEDEAKKAAEDAKKAAEDA 179
   ||||| : : ||||| : : |||||

QY 33 KKKAKEAKKYAKAAKAEKYEYAAAEAKYKAEAKAAKAEAA 74
   ||||| : : ||||| : : |||||
Db 180 KKKAEEAEKKKA-AAEAANKKAAVEAAKKKAAAAAAAARKAA 220
   ||||| : : ||||| : : |||||

RESULT 3
TOLA_HABIN STANDARD; PRT; 372 AA.
AC P44678; P94810;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tola protein.
DE TOLA OR HI0383.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
[1] RP SEQUENCE FROM N.A.
```





RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RT "HIFPM-A target for differentiation of M.tuberculosis and M.bovis.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BCG / Tokyo;  
RA Matsunoto S., Yukitake H., Matsuo T., Minoda T., Yamada T.;  
RT "Identification of a novel protein generating bacterial slow growth  
from Mycobacterium";  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Y18421; CAB46493.1; -.  
DR EMBL; AB013441; BAA78330.1; -.  
DR HSSP; P02346; 1HUU.  
DR InterPro; IPR00119; Bac.DNABind.  
DR InterPro; IPR001386; Histone\_H1/H5.  
DR Pfam; PF00216; Bac.DNA binding; 1.  
DR PRINTS; PR00624; HISTONEH5.  
DR ProDom; PD000945; Bac.DNABH5.  
DR SMART; SM00411; BHL; 1.  
DR PROSITE; PS00045; HISTONE LIKE; 1.  
DR DNA-binding; DNA condensation; Repeat.  
KW DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.  
FT CONFLICT 199 199 A -> T (IN REF. 2).  
SQ SEQUENCE 205 AA; 21262 MW; 19FCE67885DFE6A8 CRC64;  
  
Query Match 35.5%; Score 130; DB 1; Length 205;  
Best Local Similarity 49.4%; Pred. No. 0.0019;  
Matches 42; Conservative 7; Mismatches 26; Indels 10; Gaps 3;  
  
Qy 1 AKYAKKEKAYAKKAEKAAKKA--EAKYKAA-----EAKKAKAEAKKYAKAAEKK 52  
Db 111 AKYAK--KAPAKATKAAKAAKATKAPAKAKATKAPAKAVATKSPAKVTKAVKTAV 168  
  
Qy 53 EYAAAEAKYKAEAAKAAKAAVEA 77  
Db 169 KASVKKATKAPAKKAAAKRPATKA 193  
  
RESULT 6  
ID MST2 DROHY STANDARD; PRT; 1391 AA.  
AC Q08696;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Axoneme-associated protein mst101(2).  
OS MST101(2).  
GN Drosophila hydei (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7224;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=95045538; PubMed=7957199;  
RA Neesen J., Padmanabhan S., Buenemann H.;  
RT "Randomly arranged repeats of a novel highly charged 16-amino-acid  
motif representing the major component of the sperm-tail-specific

RT axoneme-associated protein family Dhmat101 form extended  
RT alpha-helical rods within the extremely elongated spermatozoa of  
RT Drosophila hydei";  
RL Eur. J. Biochem. 225:1089-1095(1994).  
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY  
SPERMATIDS.  
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
CC -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT  
STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM  
REPEATS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X73481; CAA51876.1; -.  
DR PIR; S34154; S34154.  
DR FlyBase; FBgn0020733; Dhyd\mst101(2).  
KW Sperm; Repeat; Multigene family; Polymorphism.  
FT DOMAIN 332 1268  
[KR] -K-X-C-X-X-A-K-X-A-K-X-X-K-X-X-X-X-E.  
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;  
  
Query Match 35.5%; Score 130; DB 1; Length 1391;  
Best Local Similarity 51.7%; Pred. No. 0.0087;  
Matches 45; Conservative 3; Mismatches 23; Indels 16; Gaps 5;  
  
Qy 2 KKYAKKEK--AYAKKAEKAAK---AEAKYKAAEKKAKAEAKKYAKAAK-----AE 50  
Db 512 KKLAKKEKETAEKKCKEAKKRRKKEAAEKKCKEAKKRAAEKKCKEAKKCKEAKKRAEAE 571  
  
Qy 51 KK--EYAAAEAKYKAEAAK---AAAKE 72  
Db 572 KKKCKEAKKERKAEAAEKKCEAAKKE 598  
  
RESULT 7  
DBH MYCTU STANDARD; PRT; 214 AA.  
AC P95T09;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp) (21-kDa  
laminin-2-binding protein).  
GN HUP OR HLP OR LBP21 OR RV2986C OR MT3064 OR MTCV349.01.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.

```

RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE OF 71-86, AND DNA-BINDING.
RC STRAIN=H37Rv;
RA Prasad H.K., Annapurna P.S., Dey A.B., Tyagi J.S., Jain N.K.,
RA Savita P.;
RL Submitted (DEC-1997) to the SWISS-PROT data bank.
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z83018; CAB05427.1; -.
DR EMBL; AB007127; AAK47393.1; -.
DR HSSP; P02346; 1HUU.
DR TIGR; MT3064; -.
DR TuberculList; RV2986c; -.
DR InterPro; IPR000119; Bac DNABind.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00216; Bac DNA binding; 1.
DR PRINTS; PD00624; HISTONEHS.
DR ProDom; PD000945; Bac DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
DR DNA-binding; DNA condensation; Repeat.
KW DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 1 90
FT DOMAIN 101 214 DEGENERATE REPEATS REGION.
SQ SEQUENCE 214 AA; 22187 MW; CB09AF20FB353544 CRC64;

Query Match 35.4%; Score 129.5; DB 1; Length 214;
Best Local Similarity 45.7%; Pred. No. 0.0021;
Matches 43; Conservative 4; Mismatches 28; Indels 19; Gaps 3;

QY 1 AKYAKKEKAYAKKAEKAAKAEAK-----AYKAAEAKKAKAEKAYAKKAEKAEK 54
Db 111 AKYAK--KAPAKATKAAKAAATKAPAKATKAPAKATKAPAKATKAPAKATKAPAKV 169
QY 55 AAA-----EAKYKAEAKAAKAEKAEKAEK 77
Db 169 TKAVKKTAVKAVSRKATKAPAKKAAKAAKRPATKA 202

RESULT 8
DBH_MYCSM
ID DBH_MYCSM STANDARD; PRT; 208 AA.
AC Q92HC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).
GN HUP OR HLP.
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1772;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 700084 / mc(2)155;
RX MEDLINE=99110209; PubMed=9894918;
RA Lee B.H., Murugasu-Oei B., Dick T.;
RT "Upregulation of a histone-like protein in dormant Mycobacterium
RT smegmatis.";
RL Mol. Gen. Genet. 260:475-479(1998).
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF068138; AAD13809.1; -.
DR HSSP; P02346; 1HUU.
DR InterPro; IPR000119; Bac DNABind.
DR Pfam; PF00216; Bac DNA binding; 1.
DR ProDom; PD000945; Bac DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
DR DNA-binding; DNA condensation; Repeat.
KW DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 1 90
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
SQ SEQUENCE 208 AA; 21230 MW; CA5F577F61F7EF09 CRC64;

Query Match 35.2%; Score 129; DB 1; Length 208;
Best Local Similarity 52.4%; Pred. No. 0.0023;
Matches 44; Conservative 6; Mismatches 24; Indels 10; Gaps 4;

QY 1 AKYAKK---EKAYAKK--AEKAAKAEAK--AYKAAEAKKAKAEKAYAKKAEKAEK 53
Db 111 AKKAAKAPAKKAAKAAKKTATKAAKAPAKKAAKATKAPAKKAAKATKAPAKKAAKAPAKK-- 168
QY 54 YAAAEKVKKAEKAAKAEKAEKAEKAEK 77
Db 169 -AATKAPAKKAAKAAKAPAKKAAKATKA 191

RESULT 9
MST1_DROHY
ID MST1_DROHY STANDARD; PRT; 344 AA.
AC Q08695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mst101(1).
GN MST101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7224;
RN [1]
RC SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=94200512; PubMed=8150205;
RA Neesen J., Buenemann H., Heinlein U.A.;
RT "The Drosophila hydei gene Dmst101(1) encodes a testis-specific,
RT repetitive, axoneme-associated protein with differential abundance in
RT Y chromosomal deletion mutant flies.";
RL Dev. Biol. 162:414-425(1994).
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC ASSOCIATED WITH AXONEMAL STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: Testis. LOCATED IN SPERMATOCYTES AND

```

```

CC SPERMATID BUNDLES.
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73480; CAAS1875.1; -.
CC
CC DR PIR; S34153.
CC DR FlyBase; FBgn011816; Dhyd\mat101(1).
CC KW Sperm; Repeat; Multigene family.
CC FT DOMAIN 58 337
CC FT 19 X 16 AA APPROXIMATE TANDEM REPEATS OF
CC FT K-K-C-X-X-A-[KQ]-K-X-X-E-X-A-X.
CC FT SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;
CC
CC Query Match 34.8%; Score 127.5; DB 1; Length 344;
CC Best Local Similarity 46.9%; Pred. No. 0.0043;
CC Matches 46; Conservative 6; Mismatches 23; Indels 23; Gaps 5;
CC
CC Qy 1 AKYAKKEKAYAKAEAKAE--AKAYKAAEAATKAK--AEAKYAKAEAKAE----- 50
CC Db 81 AKK--EKAAEKKKCAEAANKKEAEAKKKCAEAANKKEAEAKKKCAEAANKKEAEAK 138
CC
CC Qy 51 -----KKEVAAAEAKYKAEAAK-----AAKAEAAEYEA 77
CC Db 139 KKCAEAANKKEAEAKKKCAEAANKKEAEAKKKCAEA 176
CC
CC RESULT 10
CC HIG_STRPU STANDARD; PRT; 217 AA.
CC AC P07796;
CC DT 01-AUG-1988 (Rel. 08, Created)
CC DT 01-AUG-1988 (Rel. 08, Last sequence update)
CC DT 15-JUN-1999 (Rel. 38, Last annotation update)
CC DE Histone H1-gamma, late.
CC OS Strongylocentrotus purpuratus (Purple sea urchin).
CC OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
CC OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
CC OC Strongylocentrotus.
CC OX NCBI_TaxID=7668;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=87172742; PubMed=3031476;
CC RA Knowles J.A., Lai Z.-C., Childs G.J.;
CC RT "Isolation, characterization, and expression of the gene encoding the
CC late histone subtype H1-gamma of the sea urchin Strongylocentrotus
CC purpuratus."
CC RL Mol. Cell. Biol. 7:478-485 (1987).
CC CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC CC -!- SUBCELLULAR LOCATION: Nuclear.
CC CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M16033; AAA30059.1; -.
CC DR PIR; A26721; A26721.
CC DR HSP; P02259; IHST.
CC DR InterPro; IPR001386; Histone H1/H5.
CC DR Pfam; P00538; linker_histone; I.
CC
CC Qy 2 KKYAKKEKAYAKAEAK-----AAKAEAKYKAEAKKKAKAEAKYKAEAK 51
CC Db 96 KSDAQAPDAKAAKAAKAAKKEKKAARSKAKKEKLAAKKAAKKTTKYKVPAAKAK 155
CC
CC Qy 52 KEVAAAEAKYKAEAKAAKAEAA 74
CC Db 52 KEVAAAEAKYKAEAKAAKAEAA 74

```

```

DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
DR SQ SEQUENCE 217 AA; 22658 MW; C7251ED3413B185 CRC64;
CC
CC Query Match 34.7%; Score 127; DB 1; Length 217;
CC Best Local Similarity 45.2%; Pred. No. 0.0033;
CC Matches 38; Conservative 5; Mismatches 29; Indels 12; Gaps 1;
CC
CC Qy 2 KKYAKKEKAYAKAEK-----AAKAEAKYKAEAKKKAKAEAKYKAEAKA 49
CC Db 125 KTKARKEKVAAKAAKATKTKTKVKKPAKPAKPAKPAKPAKPAKPAKPAKPAK 184
CC
CC Qy 50 EKKEVAAAEAKYKAEAKAAKAEAA 73
CC Db 185 PAKAAAKPAKAAKPAKPAKPA 208
CC
CC RESULT 11
CC H1B_STRPU STANDARD; PRT; 211 AA.
CC AC P15369;
CC DT 01-APR-1990 (Rel. 14, Created)
CC DT 01-APR-1990 (Rel. 14, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE Histone H1-beta, late embryonic.
CC OS Strongylocentrotus purpuratus (Purple sea urchin).
CC OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
CC OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
CC OC Strongylocentrotus.
CC OX NCBI_TaxID=7668;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=88246461; PubMed=2837660;
CC RA Lai Z.-C., Childs G.;
CC RT "Characterization of the structure and transcriptional patterns of
CC the gene encoding the late histone subtype H1-beta of the sea urchin
CC Strongylocentrotus purpuratus."
CC RL Mol. Cell. Biol. 8:1842-1844 (1988).
CC CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC CC -!- SUBCELLULAR LOCATION: Nuclear.
CC CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20314; AAA30052.1; -.
CC DR PIR; A28100; A28100.
CC DR HSP; P02259; IHST.
CC DR InterPro; IPR001386; Histone H1/H5.
CC DR InterPro; IPR003216; Linkerhist_N.
CC DR Pfam; P00538; linker_histone; I.
CC DR ProDom; PD000373; Linkerhist_N; 1.
CC DR SMART; SM00526; H15; 1.
CC KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
CC SQ SEQUENCE 211 AA; 22169 MW; 9P214581334BBE7A CRC64;
CC
CC Query Match 34.3%; Score 125.5; DB 1; Length 211;
CC Best Local Similarity 49.4%; Pred. No. 0.0041;
CC Matches 41; Conservative 6; Mismatches 25; Indels 11; Gaps 3;
CC
CC Qy 2 KKYAKKEKAYAKAEAKAAKAEAK-KKAKAEAKYKAEK-----AAKAEK 51
CC Db 96 KSDAQAPDAKAAKAAKAAKKEKKAARSKAKKEKLAAKKAAKKTTKYKVPAAKAK 155
CC
CC Qy 52 KEVAAAEAKYKAEAKAAKAEAA 74
CC Db 52 KEVAAAEAKYKAEAKAAKAEAA 74

```

```

RX MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Porechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides.";
RL Eur. J. Biochem. 104:559-566(1980).
RN [2]
RP SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.A.
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Porechinus angulosus. 2. Sequence of the C-terminal CNBR peptide and
RT the entire primary structure.";
RL Eur. J. Biochem. 104:567-578(1980).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR; A02586; HSURIP.
DR HSP; P02259; LHST.
DR InterPro; IPR001386; Histone_H1/H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
FW Sperm.
FT VARIANT 144 144 K -> R.
FT SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match 33.6%; Score 123; DB 1; Length 248;
Best Local Similarity 48.7%; Pred. No. 0.0071;
Matches 37; Conservative 9; Mismatches 28; Indels 2; Gaps 20

QY 1 AKKYAKKEKAYAKAEKAAK-KAEAKYKAAEKAKKAKAEKAKYA-KAAAEKKEKYEAAA 58
Db 145 AKAAAKKKAALAKKAAAKKAAAKKAAAKKAAKAKKPKKAAKAKKPAKSPKAKKPAKSP 200
QY 59 AKYKAEAAKAAKAEAA 74
Db 205 KKKKAKRSPKAKKAA 220

RESULT 14
RS6_AEDAE RS6_AEDAE STANDARD; PRT; 346 AA.
ID AC Q9U761;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S6.
GN RPS6.
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Aedes.
OX NCBI_TaxID=7159;
RN [1]
RP SEQUENCE FROM N.A.
RA Gavino V.H., Fallon A.M.;
RL "Aedes mosquitoes ribosomal protein S6 cDNA.";
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CONTROLLING CELL GROWTH
CC AND PROLIFERATION THROUGH THE SELECTIVE TRANSLATION OF PARTICULAR
CC CLASSES OF MRNA (By similarity).
CC -1- PTM: Ribosomal protein S6 is the major substrate of protein
CC kinases in eukaryote ribosomes (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 40.7743 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-5  
Perfect score: 366  
Sequence: 1 AKKYAKKEKAYAKAEKAAK.....EAKYKAEAKAAKAEAYEA 77

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mmc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_todent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	40.4	394	16	Q8X965 escherichia
2	148	40.4	407	16	Q8ZQ06 salmonella
3	146	39.9	376	16	Q8Z8C1 salmonella
4	145.5	39.8	395	2	Q937K4 erwinia chr
5	142.5	38.9	372	2	Q9WXX1 pseudomonas
6	140	38.3	389	16	Q9CM70 pasteurella
7	138	37.7	232	10	Q39576 chlamydomon
8	138	37.7	1701	5	O61164 plasmodium
9	135.5	37.0	1341	16	Q98KG7 rhizobium 1
10	134	36.6	1128	5	Q26947 trypanosoma
11	130.5	35.7	388	16	Q8ZG22 yersinia pe
12	128	35.0	200	16	Q8XVW7 talstonia s
13	125.5	34.3	212	3	Q93946 candida alb
14	124.5	34.0	1866	5	Q8T5C8 plasmodium
15	122.5	33.5	311	12	Q84528 parametium
16	122	33.3	243	16	Q92A67 listeria in

17	122	33.3	755	10	Q9FP71 oryza sativ
18	121.5	33.2	239	16	Q8Y5W4 listeria mo
19	120.5	32.9	284	10	O65794 triticum ae
20	119.5	32.7	191	5	O46142 mytilus edu
21	118.5	32.4	191	5	O46141 mytilus edu
22	118.5	32.4	191	5	O46143 mytilus edu
23	118.5	32.4	191	5	O46362 mytilus edu
24	118.5	32.4	233	5	O18319 chironomus
25	118.5	32.4	275	5	O01395 drosophila
26	118	32.2	969	5	Q9NDI9 plasmodium
27	117.5	32.1	265	10	Q39598 chlamydomon
28	117.5	32.1	288	10	O65795 triticum ae
29	117	32.0	467	4	Q9H9F1 homo sapien
30	116.5	31.8	1671	5	Q8T5C9 plasmodium
31	116	31.7	224	13	Q90ZD7 bufo bufo g
32	116	31.7	275	10	Q9XHL9 triticum ae
33	115.5	31.6	98	5	Q8WQ44 leishmania
34	115.5	31.6	111	5	Q8T9R3 leishmania
35	115.5	31.6	191	5	O46140 mytilus edu
36	115.5	31.6	243	5	Q23784 chironomus
37	115.5	31.6	485	10	Q8RXD0 arabidopsis
38	115.5	31.6	924	10	Q9SU08 arabidopsis
39	114	31.1	68	5	Q26907 trypanosoma
40	114	31.1	198	4	Q9H8H4 homo sapien
41	114	31.1	237	10	Q9SWU2 triticum ae
42	114	31.1	581	16	Q9RU45 deinococcus
43	114	31.1	607	5	Q9V6S7 drosophila
44	113.5	31.0	845	5	Q9Y1P8 plasmodium
45	113	30.9	236	10	Q9SWU3 triticum ae

## ALIGNMENTS

### RESULT 1

Q8X965 PRELIMINARY; PRT; 394 AA.  
ID Q8X965  
AC Q8X965;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Membrane spanning protein, required for outer membrane integrity  
DE (Membrane spanning protein Tola).  
GN TOLA OR Z0907 OR ECS0774.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_taxid=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RN Nature 409:529-533(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe I.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
DR EMBL; AF005252; AAG55075.1; -.  
DR EMBL; AP002553; BAB34197.1; -.







Q9WWX1 PRELIMINARY; PRT; 372 AA.  
ID Q9WWX1  
AC Q9WWX1  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE TOLA protein.  
GN TOLA.  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT-2;  
RA MEDLINE=961198174; PubMed=8626299;  
RX Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;  
RT "The Pseudomonas putida peptidoglycan-associated outer membrane  
lipoprotein (PAL) is involved in maintenance of the integrity of the  
cell envelope.";  
RN J. Bacteriol. 178:1699-1706(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT-2;  
RA Ramos-Conzalez I.;  
RN Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT-2;  
RA Rodriguez-Herva J.J.;  
RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT-2;  
RA MEDLINE=96422022; PubMed=8824639;  
RX Rodriguez-Herva J.J., Ramos J.;  
RT "Characterization of an OprL null mutant of Pseudomonas putida.";  
RN J. Bacteriol. 178:5836-5840(1996).  
DR EMBL; X74218; CAB50780.1; -;  
DR InterPro; IPR001386; Histone\_H1/H5.  
DR PRINTS; PR00624; HISTONEH5.  
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;  
Query Match 38.9%; Score 142.5; DB 2; Length 372;  
Best Local Similarity 46.9%; Pred. No. 0.0043;  
Matches 46; Conservative 10; Mismatches 21; Indels 21; Gaps 4;  
Qy 1 AKYAKKEKA-----YAKKEKAAKKA---EAKYKAAEAKKKAKAEAKYA--- 44  
Db 148 AKKAAEKQADIAKKAEDEAKKAEAKKAAAEAKKAAEDAKKAAEAKKAAED 207  
Qy 45 ---KAAKAEKKYAAAEAKYK---AEAAKAAAEKAAAYEA 77  
Db 208 AKKAAAEADAKKAAAEAKKAAADAKKAAAEAKKAAAEAKKAAAEAKKAAAE 245  
RESULT 6  
Q9CM70 PRELIMINARY; PRT; 389 AA.  
ID Q9CM70  
AC Q9CM70  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE TOLA.  
GN TOLA OR PM0968.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida PM70.";  
Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
DR EMBL; AE006136; AAK03052.1; -;  
DR HSSP; P19934; 1TOL.  
DR InterPro; IPR000533; Tropomyosin.  
DR PRINTS; PR00194; TROPOMYOSIN.  
KW Complete proteome.  
SQ SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9E94B CRC64;  
Query Match 38.3%; Score 140; DB 16; Length 389;  
Best Local Similarity 50.5%; Pred. No. 0.0007;  
Matches 46; Conservative 10; Mismatches 21; Indels 14; Gaps 4;  
Qy 1 AKKYA-----KKEKAYAKKAEAKKAE- AKAYKAAEA---KKKAKAEAKKYAKAAKAE- 50  
Db 167 AKRLAALAKQAEAEAKAKAEAEAKKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAE 226  
Qy 51 ----KKEVAAAEAKYKAEAAKAAAEKAAAYEA 77  
Db 227 EAKAKAEKAEAKAEAKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 257  
RESULT 7  
Q39576 PRELIMINARY; PRT; 232 AA.  
ID Q39576  
AC Q39576;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Histone H1.  
GN CHI.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96120862; PubMed=8590479;  
RA Fabry S., Muller K., Lindauer A., Park P.B., Cornelius T., Schmitt R.;  
RT "The organization structure and regulatory elements of Chlamydomonas  
histone genes reveal features linking plant and animal genes.";  
RL Curr. Genet. 28:333-345(1995).  
DR EMBL; U16726; AAA98452.1; -;  
DR HSSP; P02259; 1HST.  
DR InterPro; IPR001386; Histone H1/H5.  
DR InterPro; IPR003216; Linkerhist N.  
DR Pfam; PF00538; linker\_histone; 1.  
DR ProDom; PD000373; Linkerhist\_N; 1.  
DR SMART; SM00526; H15; 1.  
SQ SEQUENCE 232 AA; 24693 MW; 2D006AE44A8FA037 CRC64;  
Query Match 37.7%; Score 138; DB 10; Length 232;  
Best Local Similarity 49.4%; Pred. No. 0.00061;  
Matches 41; Conservative 4; Mismatches 28; Indels 10; Gaps 2;  
Qy 1 AKYAKKEKAYAKKAEAKKAEAE-----KAYKAAEAKKAEAKKAEAKKAEAKKAE 51  
Db 103 AKKAAATPKAAAPKKEGAVKTKAPKKEGKPKSAKAEKKPKKEGKKAAPKAEK 162  
Qy 52 KEKAAAEAKYKAEAAKAAAEKAA 74  
Db 163 KPKAAKPAK-KTPTKAAAKPKA 184  
RESULT 8  
O61164 PRELIMINARY; PRT; 1701 AA.  
ID O61164  
AC O61164;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Erythrocyte binding protein.

```
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=72329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235(1998).
DR EMBL; AF031886; AAC0366.1; -
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87CE8A CRC64;

Query Match 37.7%; Score 138; DB 5; Length 1701;
Best Local Similarity 54.4%; Pred. No. 0.0043;
Matches 43; Conservative 8; Mismatches 16; Indels 12; Gaps 4;

Qy 5 AKKEKAVAKAEKAAKAEKAYKAAKAKKAKA-----EAKKYAKAA-KAEKKEVA--- 55
Db 1224 AKKAEAAKAEERKKAEAA-AKALERKKSEAAKALERKKAEAAKAEKKKAEAA 1282

Qy 56 --AAEAKYKAAEAAKAAKE 72
Db 1283 KKAEEKKKAAEAAKAE 1301

RESULT 9
Q98KG7 ID Q98KG7 PRELIMINARY; PRT; 1341 AA.
AC Q98KG7
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein mll1482.
GN MLL1482.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002997; BAB48847.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1341 AA; 138670 MW; 79C5D8149C969BC8 CRC64;

Query Match 37.0%; Score 135.5; DB 16; Length 1341;
Best Local Similarity 48.3%; Pred. No. 0.0053;
Matches 43; Conservative 10; Mismatches 21; Indels 15; Gaps 3;

Qy 1 AKKYAKKEKAYKAAKAAK-----KAEKAYKAAEAKKAKAEKAYKAA--- 46
Db 1177 AEELRQEEAARLKAEADAKAKAEAEQAQAEAKAEKAAQAEAKADAEAKAEAKAK 1236

Qy 47 AKAEKKEYYAAAEAKYKAAE-AKAAKAEAA 74
Db 1237 ADAEAKADAEAKADADAKAAAEQQA 1265

RESULT 10
Q26947 ID Q26947 PRELIMINARY; PRT; 1128 AA.
```

```
AC Q26947;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE R27-2 protein.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]_TaxID=5693;
RP SEQUENCE FROM N.A.
RC STRAIN=SYLVIO X-10;
RX MEDLINE=931165082; PubMed=8381919;
RA Otsu K., Donelson J.E., Kirchhoff L.V.;
RT "Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amino acid repeats by targeted insertion of the neomycin phosphotransferase gene.";
RL Mol. Biochem. Parasitol. 57:317-330(1993).
DR EMBL; L04603; AAA96494.1; -
DR InterPro; IPR004089; Chmtaxis transd.
DR InterPro; IPR001230; Prenyl site.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN 1.
SQ SEQUENCE 1128 AA; 119686 MW; 7997745D32B83656 CRC64;

Query Match 36.6%; Score 134; DB 5; Length 1128;
Best Local Similarity 52.0%; Pred. No. 0.0059;
Matches 39; Conservative 9; Mismatches 23; Indels 4; Gaps 2;

Qy 1 AKKYAKKEKAYKAAKAAKAEKAYKAAEAKKKAKAEAKKAKAAK-AEKKEYAAAEAA 59
Db 975 ATKVAEAEK---QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 1031

Qy 60 KYKAEAAKAAKAEAA 74
Db 1032 TKVAEAEKQKAAEAA 1046

RESULT 11
Q8ZGZ2 ID Q8ZGZ2 PRELIMINARY; PRT; 388 AA.
AC Q8ZGZ2
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tola colicin import membrane protein.
GN TOLA OR YPO1123.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]_TaxID=632;
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell I., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414145; CAC89966.1; -
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 388 AA; 40424 MW; 814447B04B30A7E7C CRC64;

Query Match 35.7%; Score 130.5; DB 16; Length 388;
Best Local Similarity 51.4%; Pred. No. 0.0039;
Matches 37; Conservative 10; Mismatches 20; Indels 5; Gaps 2;

Qy 7 KEKAVAKAEKAAKAAKAEKAYK---AABAKKKAKAEAKKAYKAAKAEKKEYAAAEAKYK 62
```



```

RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., King W.,
RA Lisee A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PCV-1 encodes a functional homosperrmidine
RT synthase.";
RL Virology 263:254-262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT PCV-1.";
RL Virology 276:27-36(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42580; AAC96576.1; -;
DR InterPro; IPR02048; EF-hand.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
SQ SEQUENCE 311 AA; 35390 MW; 97B0E9E82AFEF88A CRC64;

Query Match 33.5%; Score 122.5; DB 12; Length 311;
Best Local Similarity 48.6%; Pred. No. 0.013;
Matches 36; Conservative 9; Mismatches 26; Indels 3; Gaps 2;

Qy 1 AKKYAKKEKAYAKK-AEKAAYKAAKAYKAAEAK--KKAAEAKKYAKAAKAEKKEAYAA 57
Db 66 AEKAAEKERVKAEEAAREAEKERVKAEEKAEKAEKAEKAEKAEKAEKAEKAEKAEK 125
Qy 58 EAKYKAEAAKAAK 71
Db 126 EKAPAREAKAAK 139
```

Search completed: March 10, 2003, 12:25:57  
Job time : 41.7743 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 40.1245 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-5

Perfect score: 366

Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYAKAEKAAKAEKAEYEA 77

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	366	100.0	77	21	Copolymer molecule
2	351.5	96.0	86	21	Copolymer molecule
3	289	79.0	109	21	Copolymer molecule
4	282.5	77.2	66	21	Copolymer molecule
5	245.5	67.1	56	21	Copolymer molecule
6	144.5	39.5	100	21	Peptide #10 used i
7	144.5	39.5	100	21	Amino acid polymer
8	144.5	39.5	100	22	Poly-Lys-Ala used
9	144.5	39.5	100	22	Nucleic acid trans
10	142.5	38.9	154	11	Recombinant copoly

11	140	38.3	106	11	AA06446	Recombinant copoly
12	134	36.6	643	16	AA084568	Trypanosoma cruzi
13	131	35.8	45	21	AA082572	Copolymer molecule
14	130	35.5	205	16	AA020575	Mycobacterium bovi
15	130	35.5	472	16	AA084569	Trypanosoma cruzi
16	130	35.5	564	16	AA084565	Trypanosoma cruzi
17	129.5	35.4	214	20	AA034055	M. tuberculosis hi
18	129.5	35.4	214	21	AA057353	M. tuberculosis hi
19	123	33.6	433	18	AA030256	Zuotin. Saccharom
20	123	33.6	433	20	AA032954	Human zuotin prote
21	123	33.6	433	22	AA070739	S cerevisiae apopt
22	122.5	33.5	140	13	AA025206	Synthetic helical
23	121.5	33.2	239	23	AB049123	Listeria monocytog
24	121	33.1	223	20	AA014928	Amino acid sequenc
25	117	32.0	279	22	AA030592	Human DNA modifca
26	117	32.0	467	22	AA094309	Human protein sequ
27	116	31.7	427	22	AA070868	C albicans apoptos
28	114.5	31.3	219	21	AA076981	Quail H1 histone p
29	114	31.1	198	22	AA095499	Human protein sequ
30	114	31.1	607	22	AB062173	Drosophila melanog
31	114	31.1	1507	21	AA024128	Plasmodium yoelii
32	112.5	30.7	146	22	AA025508	Human protein sequ
33	112	30.6	222	18	AA029476	Human histone H1 i
34	112	30.6	222	20	AA034033	Human histone H1 i
35	112	30.6	222	21	AA057331	Human histone H1 i
36	110.5	30.2	158	20	AA034068	Histone H1 isoform
37	110.5	30.2	158	21	AA057366	Human histone H1.5
38	110.5	30.2	226	20	AA034060	Human histone H1 i
39	110.5	30.2	226	21	AA057358	Human histone H1 i
40	110	30.1	334	22	AB028693	Novel human diagno
41	108	29.5	130	23	AA009945	Peptide fragment o
42	108	29.5	130	23	AA013237	Human linker histo
43	107.5	29.4	262	17	AA006913	T. cruzi L19E homo
44	107.5	29.4	262	20	AA032848	TcE protein sequen
45	107.5	29.4	263	20	AA023306	TcE antigen of Trp

#### ALIGNMENTS

##### RESULT 1

AA082575  
ID AA082575 standard; peptide; 77 AA.

XX AA082575;

XX 28-JUL-2000 (first entry)

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritis; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; chymotrypsin; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX WO200018794-A1.

XX 06-APR-2000.

XX 24-SEP-1999; 99WO-US22402.

XX 25-SEP-1998; 98US-0101693.

XX (YEDA ) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.

```
XX PI Gad A, Lis D;
XX PF WPI; 2000-317499/27.
XX DR
XX PT Copolymer 1 related polypeptides used as molecular weight markers for
XX PT glatiramer acetate and for treatment and prevention of immune diseases
XX PS Claim 10; Page 14; 72pp; English.
XX CC
XX CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
XX CC weight TV-marker polypeptides from the present invention. The present
XX CC invention describes polypeptides (I) for determining the molecular
XX CC weight of a copolymer (CP), which has an identified molecular weight
XX CC and an amino acid composition corresponding to the copolymer. The
XX CC polypeptides of the invention are used as molecular weight markers for
XX CC glatiramer acetate related tetrapolymers. The polypeptides may also be
XX CC used for treating and preventing immune diseases in a mammal. Autoimmune
XX CC diseases which may be treated include either cell-mediated or
XX CC demyelinating diseases and inflammatory conditions, e.g. multiple
XX CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
XX CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
XX CC disease, Guillain-Barre's syndrome, chronic immune thrombocytopaenia
XX CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
XX CC disease, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
XX CC lupus erythematosus. Mediated-mediated diseases which can be treated
XX CC include host-versus-graft disease, graft-versus-host disease, and
XX CC delayed-type hypersensitivity. The polypeptides of the invention have
XX CC defined molecular weights and physical properties which are analogous to
XX CC glatiramer acetate molecules, which makes them ideal for use as
XX CC molecular weight markers.
XX SQ Sequence 77 AA;

Query Match 100.0%; Score 366; DB 21; Length 77;
Best Local Similarity 100.0%; Pred. No. 3.2e-27;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKYAKKEKAYAKAEAKKAEAKAYKAAEAKKKAKAEAKKAYAKAEAKKKEYAAAEAK 60
DB 1 AKYAKKEKAYAKAEAKKAEAKAYKAAEAKKKAKAEAKKAYAKAEAKKKEYAAAEAK 60

QY 61 YKAEAAKAAAEKAEAYEA 77
DB 61 YKAEAAKAAAEKAEAYEA 77

RESULT 2
AAY82576
ID AAY82576 standard; peptide; 86 AA.
XX AC AAY82576;
XX XX
XX DT 28-JUL-2000 (first entry)
XX DE
XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.
XX KW Copolymer; molecular weight marker; TV-marker; immune disease;
XX KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
XX KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
XX KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
XX KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
XX KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
XX KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
XX KW Hashimoto's disease; Graves disease; Guillain-Barre's syndrome; psoriasis;
XX KW pemphigus vulgaris; systemic lupus erythematosus.
XX OS Unidentified.
XX XX
XX PN WO200018794-A1.
```

```
XX PD 06-APR-2000.
XX PF 24-SEP-1999; 99WO-US22402.
XX PR 25-SEP-1998; 98US-0101693.
XX PA (YEDA ) YEDA RES & DEV CO LTD.
XX PA (TEVA-) TEVA PHARM USA INC.
XX PI Gad A, Lis D;
XX DR WPI; 2000-317499/27.
XX PT Copolymer 1 related polypeptides used as molecular weight markers for
XX PT glatiramer acetate and for treatment and prevention of immune diseases
XX PS Claim 10; Page 14; 72pp; English.
XX CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
XX CC weight TV-marker polypeptides from the present invention. The present
XX CC invention describes polypeptides (I) for determining the molecular
XX CC weight of a copolymer (CP), which has an identified molecular weight
XX CC and an amino acid composition corresponding to the copolymer. The
XX CC polypeptides of the invention are used as molecular weight markers for
XX CC glatiramer acetate related tetrapolymers. The polypeptides may also be
XX CC used for treating and preventing immune diseases in a mammal. Autoimmune
XX CC diseases which may be treated include either cell-mediated or
XX CC antibody-mediated diseases. Such diseases include arthritic conditions,
XX CC demyelinating diseases and inflammatory conditions, e.g. multiple
XX CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
XX CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
XX CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
XX CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
XX CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
XX CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
XX CC include host-versus-graft disease, graft-versus-host disease, and
XX CC delayed-type hypersensitivity. The polypeptides of the invention have
XX CC defined molecular weights and physical properties which are analogous to
XX CC glatiramer acetate molecules, which makes them ideal for use as
XX CC molecular weight markers.
XX SQ Sequence 86 AA;

Query Match 96.0%; Score 351.5; DB 21; Length 86;
Best Local Similarity 89.5%; Pred. No. 8.1e-26;
Matches 77; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 AKYAKKEKAYAKAEAKKAEAKAYKAAEAKKKAKAEAKKAYAKAEAKKKEYAAAEAK 60
DB 1 AKYAKKEKAYAKAEAKKAEAKAYKAAEAKKKAKAEAKKAYAKAEAKKKEYAAAEAK 60

QY 61 -----YKAEAAKAAAEKAEAYEA 77
DB 61 YKAEAAKAYKAEAAKAAAEKAEAYEA 86

RESULT 3
AAY82577
ID AAY82577 standard; peptide; 109 AA.
XX AC AAY82577;
XX XX
XX DT 28-JUL-2000 (first entry)
XX DE
XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:7.
XX KW Copolymer; molecular weight marker; TV-marker; immune disease;
XX KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
XX KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
XX KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
XX KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
XX KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
XX KW Hashimoto's disease; Graves disease; Guillain-Barre's syndrome; psoriasis;
XX KW pemphigus vulgaris; systemic lupus erythematosus.
XX OS Unidentified.
XX XX
XX PN WO200018794-A1.
```

KW	inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW	Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW	diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW	Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW	pemphigus vulgaris; systemic lupus erythematosus.
XX	Unidentified.
OS	
XX	WO200018794-A1.
XX	
PD	06-APR-2000.
XX	
PF	24-SEP-1999; 99WO-US22402.
XX	
PR	25-SEP-1998; 98US-0101693.
XX	
XX	(YEDA ) YEDA RES & DEV CO LTD.
PA	(TEVA-) TEVA PHARM USA INC.
XX	
PI	Gad A, Lis D;
XX	
DR	WPI; 2000-317499/27.
XX	
PT	Copolymer 1 related polypeptides used as molecular weight markers for
PT	glatiramer acetate and for treatment and prevention of immune diseases
XX	
PS	Claim 10; Page 14; 72pp; English.
XX	
CC	AA982571 to AA982577 represent specifically claimed copolymer molecular
CC	weight TV-marker polypeptides from the present invention. The present
CC	invention describes polypeptides (1) for determining the molecular
CC	weight of a copolymer (CP), which has an identified molecular weight
CC	and an amino acid composition corresponding to the copolymer. The
CC	polypeptides of the invention are used as molecular weight markers for
CC	glatiramer acetate related tetrapolymers. The polypeptides may also be
CC	used for treating and preventing immune diseases in a mammal. Autoimmune
CC	diseases which may be treated include either cell-mediated or
CC	antibody-mediated diseases. Such diseases include arthritic conditions,
CC	demyelinating diseases and inflammatory conditions, e.g. multiple
CC	sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC	anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune
CC	uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC	purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC	disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC	myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC	lupus erythematosus. Mediated-mediated diseases which can be treated
CC	include host-versus-graft disease, graft-versus-host disease, and
CC	delayed-type hypersensitivity. The polypeptides of the invention have
CC	defined molecular weights and physical properties which are analogous to
CC	glatiramer acetate molecules, which makes them ideal for use as
CC	molecular weight markers.
XX	
SQ	Sequence 109 AA;
	Query Match 79.0%; Score 289; DB 21; Length 109;
	Best Local Similarity 67.0%; Pred. No. 6.5e-20;
	Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps
Qy	1 AKYAKK-EKAYAKKA-----EKAAKAEAKAYAKAEAKKA----- 36
	1 AKYAKKAEKAYAKKAAREKAYAKAEAKKAYAKAEAKKAYAKAEAKKA 60
Db	
Qy	37 -KAEAKKYAKAAEKKEYAAAEAK-----YKAEAAKAAKAEAYEA 77
Db	61 YKAEAKKYAKAAEKKEYAAAEAKKAEAAKAYAKAEAAKAAKAEAYEA 109
RESULT 4	
ID	AA982574
XX	ID AA982574 standard; peptide; 66 AA.
AC	
XX	AA982574;
XX	









Db 72 AKYKKAEAEYKKAKAAAEAEYKKEAEAEYKVKYKKAKAKYK 117

## RESULT 11

AAR06446  
ID AAR06446 standard; protein; 106 AA.

AC AAR06446;

XX 03-JAN-1991 (first entry)

XX Recombinant copolymer 1-19, myelin basic protein analogue.

XX Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;

KW immunological activity; autoimmune encephalomyelitis;

KW multiple sclerosis;

XX Synthetic.

XX EP383620-A.

XX 22-AUG-1990.

XX 16-FEB-1990; 90EP-0301700.

XX 07-FEB-1990; 90US-0473845.

PR 17-FEB-1989; 89US-0312541.

XX (REPL-) REPLIGEN CORP.

XX Cook KS;

XX WPI; 1990-255848/34.

DR N-PSDB; AAQ06446.

XX Producing genes encoding random polymers of aminoacid(s) - for

PT producing recombinant polypeptide(s) with biological and/or

PT immunological activity

PS Disclosure; Fig 12; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes

CC coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaN

CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to

CC express Protein A. The resulting plasmids encode fusion proteins

CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.

CC A methionine residue occurs between the Protein A and rCOP-1

CC sequences, originating from the 5' linker sequence, in order that

CC the COP-1 polypeptide may be cleaved from the fusion protein.

CC rCOP-1-19 contains oligonucleotide duplexes encoding the following

CC segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA KEA, and AAA. The

CC N-terminal alanine residue is left behind following CNBr cleavage of the

CC fusion protein.

CC The product prevents or arrests experimental autoimmune

CC encephalomyelitis. They are used to prevent, arrest or control a

CC demyelinating disorder, e.g. multiple sclerosis. They may also

CC be used as additives to hair care products to confer beneficial

CC effects on damaged hair or as supplements for diets deficient in

CC certain amino acids.

CC See also AAQ05664.

XX SQ Sequence 106 AA;

Query Match 38.3%; Score 140; DB 11; Length 106;

Best Local Similarity 54.5%; Pred. No. 4.2e-06;

Matches 42; Conservative 7; Mismatches 22; Indels 6; Gaps 3;

Qy 1 AKYKKAEKAYAKAEKAAKAEKAYKAAAEKAKAEKAYKAAAEKAEKAYKAAAEK 60

Db 25 AKKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEK 79

Qy 61 YKAEAKAAAK-EAAEY 76

||| ||||| :|||:

Db 80 KKAEEKAAAEKAKAAAYK 96

## RESULT 12

AAR84568  
ID AAR84568 standard; Protein; 643 AA.

AC AAR84568;

XX 09-MAY-1996 (first entry)

XX Trypanosoma cruzi TCR27 polypeptide, Ag15.

XX TCR27; Chagas disease; repeat unit; diagnosis; blood screening;

KW recombinant; fusion protein; glutathione-S-transferase.

XX Trypanosoma cruzi.

XX Key Location/Qualifiers

FT 329..552

FT /label= repeat\_region

FT /note= "16 of 69 repeat units of 14 amino acids"

XX W09525797-A1.

XX 28-SEP-1995.

XX 20-MAR-1995; 95WO-US03191.

XX 24-MAR-1994; 94US-0216894.

XX (KIRC/) KIRCHHOFF L V.

PA (OTSU/) OTSU K.

XX Kirchhoff LV, Otsu K;

XX WPI; 1995-344618/44.

DR N-PSDB; AAT05332.

XX New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as

PT immunoassay reagent for specific diagnosis of Chagas disease, also

PT related nucleic acid and transformed cells

XX Disclosure; Page 40-41; 68pp; English.

XX AAR84565-R84569 are polypeptides of the TCR27 protein of T. cruzi

CC The proteins are all fusion products with glutathione-S-transferase

CC (GST) and some contain a linker sequence. The TCR27 protein comprises

CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly

CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence

CC encodes the GST sequence, the Ag44 polypeptide contg. 16 of the 69

CC repeat units and also contains the amino and carboxy terminal

CC peptides of TCR27. The TCR27 polypeptides of the invention are useful

CC for the diagnosis of Chagas disease (American Trypanosomiasis), they

CC are capable of detecting anti-T. cruzi antibodies; or for blood

CC screening. The TCR27 protein has epitopes to which most T. cruzi

CC infected individuals have antibodies. The TCR27 polypeptides will not

CC react with serum from patients with leishmaniasis, schistosomiasis,

CC or autoimmune disease and are hence less likely to cause false

CC positives in diagnosis.

XX SQ Sequence 643 AA;

Query Match 36.6%; Score 134; DB 16; Length 643;

Best Local Similarity 52.0%; Pred. No. 0.0001;

Matches 39; Conservative 9; Mismatches 23; Indels 4; Gaps 2;

Qy 1 AKYKKAEKAYAKAEKAAKAEKAYKAAAEKAKAEKAYKAAAEKAEKAYKAAAEK 59

Db 490 ATKVAEAEK---QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 546

Qy 60 KYKAEAAKAAAEKAA 74

||| ||||| :|||:

Db 547 TKVAEAEKOKAEEA 561

RESULT 13  
RAY82572  
ID AAY82572 standard; peptide; 45 AA.  
XX  
AC AAY82572;  
XX  
XX 28-JUL-2000 (first entry)  
XX  
XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:2.  
XX  
XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyronimetic; haemostatic; antipsoriatic; dermatological;  
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX  
XX Unidentified.  
OS  
XX WO200018794-A1.  
PN  
XX  
XX 06-APR-2000.  
PD  
XX  
XX 24-SEP-1999; 99WO-US22402.  
PF  
XX  
XX 25-SEP-1998; 98US-0101693.  
PR  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
PA (TEVA-) TEVA PHARM USA INC.  
PA  
XX  
XX Gad A, Lis D;  
PI  
XX WPI; 2000-317499/27.  
DR  
XX  
XX Copolymer 1 related polypeptides used as molecular weight markers for  
FT glatiramer acetate and for treatment and prevention of immune diseases  
PT  
XX  
XX Claim 10; Page 14; 72pp; English.  
PS  
XX  
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular  
CC weight of a copolymer (CP), which has an identified molecular weight  
CC and an amino acid composition corresponding to the copolymer. The  
CC polypeptides of the invention are used as molecular weight markers for  
CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
CC used for treating and preventing immune diseases in a mammal. Autoimmune  
CC diseases which may be treated include either cell-mediated or  
CC antibody-mediated diseases. Such diseases include arthritic conditions,  
CC demyelinating diseases and inflammatory conditions, e.g. multiple  
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
CC lupus erythematosus. Mediated-mediated diseases which can be treated  
CC include host-versus-graft disease, graft-versus-host disease, and  
CC delayed-type hypersensitivity. The polypeptides of the invention have  
CC defined molecular weights and physical properties which are analogous to  
CC glatiramer acetate molecules, which makes them ideal for use as  
CC molecular weight markers.  
XX  
SQ Sequence 45 AA;  
Query Match 35.8%; Score 131; DB 21; Length 45;

Best Local Similarity 50.6%; Pred. No. 1.1e-05;  
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

QY 1 AKKYAKKEKAYAKAEKAKAEKAYAKAEKAKAEKAKAEKAKAEKAYAKAEKAKAEK 60  
DB 1 AKKYAKK-----AKAEKA-----KKAYKAEAKKAAKYE----- 29  
QY 61 YKAEAAKAAKAEAAVEA 77  
DB 30 -KAAAEKAAKAEAAVEA 45

RESULT 14  
AAB20575  
ID AAB20575 standard; Protein; 205 AA.  
XX  
AC AAB20575;  
XX  
DT 08-DEC-2000 (first entry)  
DE Mycobacterium bovis BCG strain MDPI protein SEQ ID NO:2.  
XX  
KW Mycobacterium bovis BCG strain; mycobacterium DNA binding protein 1;  
KW MDPI; slow growing acid-fast bacterium protein; immunogenicity;  
KW pathogenic acid-fast bacterium; diagnosis; tuberculosis; MAC;  
KW Mycobacterium avium intracellular complex; Hansen's disease.  
XX  
OS Mycobacterium bovis.  
XX  
PN WO200044905-A1.  
XX  
PD 03-AUG-2000.  
XX  
XX 28-JAN-2000; 2000WO-JP00455.  
PF  
XX 29-JAN-1999; 99JP-0022588.  
PR  
XX (SAKA ) OTSUKA PHARM CO LTD.  
PA (YAMA/) YAMADA T.  
PA (MATS/) MATSUMOTO S.  
XX  
PI Yamada T, Matsumoto S;  
XX  
DR WPI; 2000-543393/49.  
DR N-PSDB; AAA88001.  
XX  
XX New Mycobacterium DNA binding protein 1 immunogenic for pathogenic  
PT acid-fast bacteria for diagnosis, treatment and prevention of diseases  
PT such as tuberculosis  
XX  
PS Claim 1; Fig 2A; 62pp; Japanese.  
XX  
CC The present sequence represents the Mycobacterium bovis BCG strain  
CC mycobacterium DNA binding protein 1 (MDPI), which has immunogenicity  
CC against pathogenic acid-fast bacteria. The MDPI protein can be used for  
CC diagnosis and treatment of diseases involving pathogenic acid-fast  
CC bacteria, such as tuberculosis, Mycobacterium avium intracellular  
CC complex (MAC) and Hansen's disease.  
XX  
SQ Sequence 205 AA;  
Query Match 35.5%; Score 130; DB 21; Length 205;  
Best Local Similarity 49.4%; Pred. No. 7.2e-05;  
Matches 42; Conservative 7; Mismatches 26; Indels 10; Gaps 3;

QY 1 AKKYAKKEKAYAKAEKAKAEKAKAEKAKAEKAKAEKAKAEKAKAEKAKAEKAK 52  
DB 111 AKKVAK--KAPAKATKAAKAAATKAPAKAAATKAPAKAAATKAPAKAAATKAPAKAAATKAVKTAV 168  
QY 53 EYAAAEAKYKAEAAKAEAAVEA 77  
DB 169 KASVRKAAATKAPAKAAKAAKRPATKA 193



THIS PAGE BLANK (USPTO)







[illegible]



```
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36182
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005529.7
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
; OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUE 2.00e-33
; OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALUE 3.00e-22
US-09-864-761-36182

Query Match      28.6%; Score 104.5; DB 10; Length 617;
Best Local Similarity 40.0%; Pred. No. 0.031;
Matches 32; Conservative 18; Mismatches 23; Indels 7; Gaps 3;

Qy 2 KKYAK---KEKAYAKAEKAAKAEKAYKAEKAAKAEKAYKAKA---AKAEKKEYA 55
Db 110 EKEAKSPVKEEAKSPAEAKSPKEAKSPAEVKSPEKAKSPAEKSPKEAKSPKEEA 169

Qy 56 AAEAKYKA-EAAKAAAEKAA 74
Db 170 KSPAENVKSPEKAKSPAKEA 189

RESULT 14
US-09-820-843A-95
; Sequence 95, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 95
; LENGTH: 369
; TYPE: PRT
; ORGANISM: T. pallidum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: conserved hypothetical protein
```

```
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3322751
US-09-820-843A-95

Query Match      27.9%; Score 102; DB 9; Length 369;
Best Local Similarity 43.1%; Pred. No. 0.03;
Matches 31; Conservative 16; Mismatches 19; Indels 6; Gaps 4;

Qy 8 EKAYAKAEKAAKK--AEAKYKAAAEKAKKAKAEKAAKAEKAYKAEKAAAEKAYKAE 65
Db 269 BEARRKEAEARRKEAEAEARRKEAEAEARRKEAEAEARR--KEAEAEARRKE--AEAEARRKE-EA 324

Qy 66 AKAAAEKAAEYA 77
Db 325 BEARRKEAEFEA 336

RESULT 15
US-09-999-724-90
; Sequence 90, Application US/09999724
; Publication No. US20030022355A1
; GENERAL INFORMATION:
; APPLICANT: WICKHAM, THOMAS J.
; APPLICANT: KOVESDI, IMRE
; APPLICANT: BROUGHT, DOUGLAS E.
; TITLE OF INVENTION: VECTORS AND METHODS FOR GENE TRANSFER
; FILE REFERENCE: 212960
; CURRENT APPLICATION NUMBER: US/09/999,724
; CURRENT FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: US 09/101,751
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: WO 96US19150
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 08/700,846
; PRIOR FILING DATE: 1996-08-21
; PRIOR APPLICATION NUMBER: US 08/701,124
; PRIOR FILING DATE: 1996-08-21
; PRIOR APPLICATION NUMBER: US 08/563,368
; PRIOR FILING DATE: 1995-11-28
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-999-724-90

Query Match      27.0%; Score 99; DB 9; Length 102;
Best Local Similarity 40.5%; Pred. No. 0.015;
Matches 32; Conservative 11; Mismatches 20; Indels 16; Gaps 2;

Qy 12 AKKAEKAAKAE---AKAYKAAEA-----KCKAKAEAKKAYKAAKAEKKEYA 55
Db 4 AAEAEERATRAEKEKRAEAEAEAEAEAPAAQPDVEKPKQKQANAKAEAVQAAAEAEERATR 63

Qy 56 AAEAKYKAEAAKAAAEKAA 74
Db 64 ABEKRAEAEAEAEAEAPAA 82

Search completed: March 10, 2003, 12:53:44
Job time : 24.0479 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 13.6456 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-5  
Perfect score: 366  
Sequence: 1 AKYAKKEKAYAKKAERAAK.....EAKYKAEAKAAKEAYEA 77

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144.5	39.5	100	2	US-08-460-890A-64
2	144.5	39.5	100	3	US-08-167-641C-64
3	144.5	39.5	100	4	US-08-460-971A-64
4	144.5	39.5	100	4	US-08-462-040-64
5	134	36.6	643	2	US-08-216-894-8
6	134	36.6	643	4	US-09-115-746-8
7	130	35.5	472	2	US-08-216-894-10
8	130	35.5	472	4	US-09-115-746-10
9	130	35.5	564	2	US-08-216-894-2
10	130	35.5	564	4	US-09-115-746-2
11	129.5	35.4	214	3	US-09-041-889-27
12	123	33.6	433	1	US-08-346-849-2
13	123	33.6	433	2	US-08-293-284A-2
14	121	33.1	223	4	US-09-095-855-201
15	121	33.1	223	4	US-09-205-426-201
16	119.5	32.7	56	4	US-08-993-008A-6
17	114	31.1	1507	3	US-08-929-329-5
18	112	30.6	222	3	US-09-041-889-3
19	112	30.6	222	3	US-08-837-058-3
20	110.5	30.2	158	3	US-09-041-889-40
21	110.5	30.2	226	3	US-09-041-889-32
22	107.5	29.4	262	1	US-08-403-379A-1
23	107.5	29.4	262	2	US-08-929-414-1
24	107.5	29.4	263	2	US-08-557-309B-51
25	107.5	29.4	263	3	US-08-834-306-51
26	107.5	29.4	263	4	US-08-993-674A-51
27	107.5	29.4	263	4	US-09-256-976-51

28 106.5 29.1 48 4 US-08-993-008A-5 Sequence 5, Appli  
29 102 27.9 218 3 US-09-041-889-4 Sequence 4, Appli  
30 102 27.9 218 3 US-08-837-058-4 Sequence 4, Appli  
31 101.5 27.7 60 1 US-08-346-849-16 Sequence 16, Appl  
32 101.5 27.7 60 2 US-08-293-284A-16 Sequence 16, Appl  
33 99.5 27.2 116 3 US-09-041-889-38 Sequence 38, Appl  
34 99 27.0 100 2 US-08-460-890A-62 Sequence 62, Appl  
35 99 27.0 100 3 US-08-167-641C-62 Sequence 62, Appl  
36 99 27.0 100 4 US-08-460-971A-62 Sequence 62, Appl  
37 99 27.0 100 4 US-08-462-040-62 Sequence 62, Appl  
38 99 27.0 102 4 US-09-101-751A-90 Sequence 90, Appl  
39 99 27.0 218 4 US-09-101-751A-48 Sequence 48, Appl  
40 99 27.0 434 1 US-08-097-830B-3 Sequence 3, Appli  
41 99 27.0 434 2 US-08-456-112B-3 Sequence 3, Appli  
42 98.5 26.9 92 4 US-09-344-529-2 Sequence 2, Appli  
43 98 26.8 219 2 US-08-557-309B-54 Sequence 54, Appl  
44 97 26.5 442 3 US-08-834-306-52 Sequence 52, Appl  
45 97 26.5 442 4 US-08-993-674A-52 Sequence 52, Appl

## ALIGNMENTS

RESULT 1  
US-08-460-890A-64  
; Sequence 64, Application US/08460890A  
; Patent No. 5994109  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FASTSEQ for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,890A  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/167,641  
; FILING DATE: December 14, 1993  
; APPLICATION NUMBER: 07/855,389  
; FILING DATE: March 20, 1992  
; APPLICATION NUMBER: PCT/US93/02725  
; FILING DATE: March 19, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 212/066  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single



```
Best Local Similarity 58.4%; Pred.No. 2.7e-07;
Matches 45; Conservative 6; Mismatches 23; Indels 3; Gaps 3;

QY 1 AKTYAK-KERAYKAEKAAKABAKAYKAAEAKKKAKAEAKTKYAKA-AKAETKEEYAAA 58
   ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61
   ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 59 AKYKAEA-AKAAAKEAA 74
   ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 62 AKAKAKAKAKAKAKAKA 78
   ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
```

RESULT 4

US-08-462-040-64  
; Sequence 64, Application US/08462040  
; Patent No. 6177554  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,040  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/167,641  
; FILING DATE: December 14, 1993  
; APPLICATION NUMBER: 07/855,389  
; FILING DATE: March 20, 1992  
; APPLICATION NUMBER: PCT/US93/02725  
; FILING DATE: March 19, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 212/078  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION: "Lys Ala" in positions 3 to 100 may be  
; OTHER INFORMATION: present or absent.

US-08-462-040-64

Query Match 39.5%; Score 144.5; DB 4; Length 100;  
Best Local Similarity 58.4%; Pred.No. 2.7e-07;  
Matches 45; Conservative 6; Mismatches 23; Indels 3; Gaps 3;

```

QY   1 AKYAK-KEKAYAKAEAAKKAABAKYKAAEKKKAAEKYAKA-AKAEKKYAAA 58
Db   2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61

QY   59 AKYKAEA-AKQAAKEAA 74
Db   62 AKAKAKAKAKAKAKAKA 78

RESULT 5
US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-8

Query Match      36.6%; Score 134; DB 2; Length 643;
Best Local Similarity 52.0%; Pred. No. 1.8e-05;
Matches 39; Conservative 9; Mismatches 23; Indels 4; Gaps

QY   1 AKYAKGKAVAKAEAAKKAABAKYKAAEKKKAAEKYAKAAK-AEKVEYAAAA 59
Db   490 ATKVAEAK--QKAAATKVAAEKKQAATAETKVAAEKKQAATAETKVAAEKKQAATA 54

QY   60 KYKAAEAAGAAAKEAA 74
Db   547 TKVAAEAKQKAAEEA 561

RESULT 6
US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
```

Query Match 39.5%; Score 144.5; DB 4; Length 100;  
Best Local Similarity 58.4%;  
Matches 45; Conservative 6; Mismatches 23; Indels 3;  
Pred. No. 2.7e-07;

```
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-8

Query Match          36.6%; Score 134; DB 4; Length 643;
Best Local Similarity 52.0%; Pred. No. 1.8e-05;
Matches 39; Conservative 9; Mismatches 23; Indels 4; Gaps 2;

QY 1 AKKYAKKEKAYAKKAAKAAKAYKAAEAKKKAKAAKAYAKAAK-AEKKEYAAAAA 59
DB 490 ATKVAEAEK---QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 546
QY 60 KYKAAEAKAAKAAEA 74
DB 547 TKVAEAEKQKAAEA 561

RESULT 7
US-08-216-894-10
; Sequence 10, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-10

Query Match          35.5%; Score 130; DB 2; Length 472;
Best Local Similarity 51.4%; Pred. No. 3.1e-05;
Matches 38; Conservative 9; Mismatches 23; Indels 4; Gaps 2;

QY 1 AKKYAKKEKAYAKKAAKAAKAYKAAEAKKKAKAAKAYAKAAK-AEKKEYAAAAA 59
DB 360 ATKVAEAEK---QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 416
QY 60 KYKAAEAKAAKAAEA 73
DB 417 TKVAEAEKQKAAEA 430

RESULT 8
US-09-115-746-10
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-10

Query Match          35.5%; Score 130; DB 4; Length 472;
Best Local Similarity 51.4%; Pred. No. 3.1e-05;
Matches 38; Conservative 9; Mismatches 23; Indels 4; Gaps 2;

QY 1 AKKYAKKEKAYAKKAAKAAKAYKAAEAKKKAKAAKAYAKAAK-AEKKEYAAAAA 59
```



[illegible]

```
/ FILING DATE: 11-APR-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Campbell, Cathryn A.
/ REGISTRATION NUMBER: 31,815
/ REFERENCE/DOCKET NUMBER: P-PM 3006
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 535-9001
/ TELEFAX: (619) 535-8949
/ INFORMATION FOR SEQ ID NO: 27:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 214 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-09-041-889-27

Query Match 35.4%; Score 129.5; DB 3; Length 214;
Best Local Similarity 45.7%; Pred. No. 1.5e-05;
Matches 43; Conservative 4; Mismatches 28; Indels 19; Gaps 3;

Qy 1 AKKYAKKEKAYAKAEAKKAEAK-----AYKAAEAKKKAKAEAKKYAKAAKAEKKEY 54
Db 111 AKKVAK--KAPAKATKAAKAAATKAPARKAATKAPAKAATKAPAKAATKAPAKAVATKSPAKV 168
Qy 55 AAA-----EAKYKAEAAKAAKAEAYEA 77
Db 169 TKAVKTAVKAVSRKATKAPAKAAKRPATKA 202

RESULT 12
US-08-849-2
; Sequence 2, Application US/08346849
; Patent No. 5670483
; GENERAL INFORMATION:
; APPLICANT: Zhang, Shuguang
; APPLICANT: Lockshin, Curtis
; APPLICANT: Rich, Alexander
; APPLICANT: Holmes, Todd
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,849
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28 DECEMBER 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 33.6%; Score 123; DB 2; Length 433;
Best Local Similarity 46.2%; Pred. No. 0.00013;
Matches 36; Conservative 10; Mismatches 28; Indels 4; Gaps 2;
```

```
/ MOLECULE TYPE: protein
US-08-346-849-2

Query Match 33.6%; Score 123; DB 1; Length 433;
Best Local Similarity 46.2%; Pred. No. 0.00013;
Matches 36; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

Qy 2 KKYAKKEKAYAKAEAKKAEAKAYKAAEAKKKAKAEAKKYAKA-AKAEK---KEYAAA 57
Db 290 KEEKKEKERRKWEREGAGARAEAEAKAKAEAEAKAEAKAESEAKANASAKADKKKAEAKA 349
Qy 58 EAKYKAEAAKAAKAEAY 75
Db 350 AKKKNKRAIRNSAKEADY 367

RESULT 13
US-08-293-284A-2
; Sequence 2, Application US/08293284A
; Patent No. 5955343
; GENERAL INFORMATION:
; APPLICANT: Holmes, Todd
; APPLICANT: Zhang, Shuguang
; APPLICANT: Rich, Alexander
; APPLICANT: Dipersio, C. Michael
; APPLICANT: Lockshin, Curtis
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/293,284A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 33.6%; Score 123; DB 2; Length 433;
Best Local Similarity 46.2%; Pred. No. 0.00013;
Matches 36; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

Qy 2 KKYAKKEKAYAKAEAKKAEAKAYKAAEAKKKAKAEAKKYAKA-AKAEK---KEYAAA 57
Db 290 KEEKKEKERRKWEREGAGARAEAEAKAKAEAEAKAEAKAESEAKANASAKADKKKAEAKA 349
Qy 58 EAKYKAEAAKAAKAEAY 75
```



THIS PAGE BLANK (USPTO)

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	164	40.1	421	2	JV0057	toIA protein - Esc
2	157.5	38.5	376	2	AG0592	toIA protein [impo
3	156.5	38.3	1701	2	T09127	probable erythrocy
4	155	37.9	394	2	F90725	membrane spanning
5	155	37.9	394	2	G85576	membrane spanning
6	154	37.7	347	2	E83525	ToIA protein PA097
7	151	36.9	210	2	A25550	histone H1 - sea u
8	149	36.4	231	2	S59589	histone H1 - Chlam
9	147.5	36.1	388	2	AC0138	ToIA colicin impor
10	147	35.9	214	2	G70673	probable hupB - My
11	144	35.2	211	2	A28100	histone H1-beta, e
12	142.5	34.8	344	2	S34153	mat101-1 protein - e
13	142	34.7	372	2	G64064	outer membrane int
14	140.5	34.4	328	2	A4993	cytosolic repetiti
15	140.5	34.4	1390	2	S51364	sperm tail-specifi
16	138	33.7	217	2	A26721	histone H1-gamma,
17	137.5	33.6	1128	2	T30296	R27-2 protein - Tr
18	135	33.0	219	2	E60110	repetitive protein
19	134	32.8	206	2	S09388	histone H1 - sea u
20	133.5	32.6	384	2	B43592	outer membrane pro
21	133.5	32.6	433	2	S25194	zuotin - yeast (sa
22	132.5	32.4	206	1	HT81R1	histone H1 - rainb
23	132	32.3	248	1	HSUR1P	histone H1, gonada
24	131	32.0	208	2	T23778	histone H1.1 - Cae
25	130.5	31.9	220	2	A28456	histone H1.10 - Ch
26	130.5	31.9	243	2	AE1689	hypothetical prote
27	129.5	31.7	924	2	T06636	probable translati
28	129.5	31.7	1403	2	T11583	histone H1.01 - Ch
29	129	31.5	218	2	A23055	histone H1.01 - Ch



A;Residues: 1-173, 'P', 174-231 <FAW>  
A;Cross-references: EMBL:U16726; NID:g571479; PIDN:AAA98452.1; PID:g571480  
C;Genetics:  
A;Introns: 62/3; 101/3  
C;Superfamily: histone H1  
C;Keywords: chromosomal protein; DNA binding, nucleosome

Query Match 36.4%; Score 149; DB 2; Length 231;  
Best Local Similarity 39.5%; Pred.No. 0.00035;  
Matches 47; Conservative 7; Mismatches 29; Indels 36; Gaps 2;

QY 1 AKKYAKKEKAYKAERAAKAE-----KAYKAAEAKKKAKAEAKKYAKAAKAEK 51  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 103 AKKAATPKKAAPKKGAVGKTKPAKPKEGEKPKSAKKAEEKPKKEGEGKKAAKPAKAEK 162  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
QY 52 KEVAA-----ABAKYKAAEAACKYKAAEAACAAAKEAA 83  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 163 KPRAAKPAKTTTCAAAKPPAKAEKPPKAAKPPKAAEPKAEKPKAAKPAKAKSA 221  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 9  
AC0138  
ToLA colicin import membrane protein [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C;Accession: AC0138  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AC0138  
A>Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-388 <KOR>  
A;Cross-references: GB:AL590842; PIDN:CAC89866.1; PID:g15979190; GSPDB:GN00175  
C;Genetics:  
A;Gene: tola

Query Match 36.1%; Score 147.5; DB 2; Length 388;  
Best Local Similarity 45.4%; Pred.No. 0.00067;  
Matches 47; Conservative 9; Mismatches 21; Indels 27; Gaps 3;

QY 7 KEKAYAKKAEEAKKAEKAYK----AAEKKKAKAEAKKYAKAAKAEKKE----- 53  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 155 KAKEEQKQETRAAQAKAEADKIYKQAQAEKQKAEAEAKKEAAVAANAANKQADADAKXAV 214  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
QY 54 -----YAAAEAKYKAEAAKK-AYKAEAAKAAKAEAA 83  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 215 EVAEKAADAAEKKAADAEBKAAAKKVAAGAAAEAKKAAAEAA 258  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 10  
G70673  
probable hupB - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: G70673  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Sqaes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: G70673  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A;Residues: 1-214 <COL>  
A;Cross-references: GB:I283018; GB:AL123456; NID:g3261671; PINO:CAN05427.1; PID:g1694845  
A;Experimental source: strain H37RV  
C;Genetics:







THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 19.5949 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-6  
Perfect score: 409  
Sequence: 1 AKKYAKKEKAYAKKAERAAK.....KKAYKAEAAKAAKEAAAYEA 86

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	164	40.1	421	1 TOLA_ECOLI	P19934 escherichia
2	154	37.7	347	1 TOLA_PSEAE	P50600 pseudomonas
3	151	36.9	210	1 HI_LYPTI	P06144 lytechinus
4	147	35.9	214	1 DBH_MYCTU	P95109 mycobacteri
5	146.5	35.8	205	1 DBH_MYCBO	Q9xb18 mycobacteri
6	142.5	34.8	344	1 MST1_DROHY	Q08695 drosophila
7	142	34.7	372	1 TOLA_HAEIN	P44678 haemophilus
8	141	34.5	211	1 H1B_STRPU	P15869 strongyloce
9	140.5	34.4	1391	1 MST2_DROHY	Q08696 drosophila
10	138	33.7	217	1 H1G_STRPU	P07796 strongyloce
11	137.5	33.6	346	1 R36_AEDAE	Q9u761 aedes aegyp
12	137	33.5	349	1 R36_AEDAL	Q9u762 aedes albop
13	136.5	33.4	208	1 DBH_MYCSM	Q2zhc5 mycobacteri
14	133.5	32.6	384	1 TMPB_TREPH	P29720 treponema p
15	133.5	32.6	433	1 ZUO1_YEAST	P32527 saccharomyc
16	132.5	32.4	206	1 HI_ONCMY	P06350 oncorhynch
17	132	32.3	248	1 HI_PARAN	P02256 paracitrus
18	131	32.0	207	1 H11_CAEEL	P10771 caenorhabdi
19	130.5	31.9	219	1 H110_CHICK	P08286 gallus gall
20	129.5	31.7	1403	1 YDF3_SCHPO	Q10475 schizosacch
21	129	31.5	218	1 H101_CHICK	P08284 gallus gall
22	127	31.1	139	1 ASR_KLEPN	Q93mh5 klebsiella
23	126.5	30.9	217	1 H1_AANPL	P09426 anas platyr
24	126.5	30.9	240	1 H12_VOLCA	Q08865 volvox cart
25	125.5	30.7	217	1 H11_CHICK	P09987 gallus gall
26	125.5	30.7	223	1 H103_CHICK	P08285 gallus gall
27	123.5	30.2	224	1 H11L_CHICK	P08287 gallus gall
28	122.5	30.0	219	1 H1B_XENLA	P06893 xenopus lae
29	121.5	29.7	237	1 H1E_CHITE	P40278 chironomus
30	121.5	29.7	238	1 H1_WHEAT	P27806 triticum ae
31	121	29.6	225	1 H15_HUMAN	P16401 homo sapien
32	120.5	29.5	771	1 CALD_CHICK	P12957 gallus gall
33	120	29.3	232	1 H1B_CHITE	P40276 chironomus

34	120	29.3	325	1	TMPE_TREPA	P19649 treponema p
35	118.5	29.0	194	1	H1_SALTR	P02254 salmo trutt
36	118	28.9	218	1	H11R_CHICK	P08288 gallus gall
37	118	28.9	221	1	H1C_CHITE	P40277 chironomus
38	118	28.9	244	1	H1O_CHITH	Q07134 chironomus
39	117	28.6	218	1	H14_HUMAN	P10412 homo sapien
40	117	28.6	233	1	H11_GLYSA	P40286 glyptotendi
41	116.5	28.5	833	1	IF2_PASMU	P57873 pasteurella
42	115.5	28.2	209	1	H1A_XENLA	P06892 xenopus lae
43	115.5	28.2	216	1	H1C1_XENLA	P15866 xenopus lae
44	115.5	28.2	233	1	H11_GLYBA	P40263 glyptotendi
45	115.5	28.2	239	1	RL23_DROME	P50887 drosophila

#### ALIGNMENTS

RESULT 1  
TOLA\_ECOLI  
ID TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE TOLA protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OC Escherichia  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JM105;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levgood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
RT their products, components of a multistep translocation system in  
RT Escherichia coli."  
RL J. Bacteriol. 171:6600-6609(1989).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map."  
RL DNA Res. 3:137-155(1996).  
[4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levgood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "Tola: a membrane protein involved in colicin uptake contains an  
RT extended helical region."  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
[5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,







```
RP SEQUENCE FROM N.A.
RC STRAIN=1479;
RX MEDLINE=97080550; PubMed=8921895;
RA Sen K., Sikkema D.J., Murphy T.F.;
RT "Isolation and characterization of the Haemophilus influenzae tolQ,
RL tolR, tolA and tolB genes.";
RT Gene 178:75-81 (1996).
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32722; AAC22041.1; -.
DR EMBL; U32470; AAC44596.1; -.
DR HSSP; P19934; 1TOL.
DR TIGR; HI0383; -.
KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 8
FT TRANSMEM 9 29
FT DOMAIN 30 372
FT PERIPLASMIC (POTENTIAL).
FT V -> A (IN STRAIN 1479).
FT K -> R (IN STRAIN 1479).
FT A -> P (IN STRAIN 1479).
FT A -> R (IN STRAIN 1479).
FT V -> A (IN STRAIN 1479).
FT D -> A (IN STRAIN 1479).
FT D -> A (IN STRAIN 1479).
FT T -> A (IN STRAIN 1479).
FT L -> F (IN STRAIN 1479).
FT I -> V (IN STRAIN 1479).
FT N -> S (IN STRAIN 1479).
FT T -> A (IN STRAIN 1479).
FT S -> P (IN STRAIN 1479).
FT SEQUENCE 372 AA; 39831 MW; 266ECF05C6C95544 CRC64;
Query Match 34.7%; Score 142; DB 1; Length 372;
Best Local Similarity 52.1%; Pred. No. 0.0008;
Matches 49; Conservative 9; Mismatches 22; Indels 14; Gaps 5;
OY 1 AKKYA-----KKEKAYAKAEAKK-----AEAKYAKAEAKKAEAKKAEAKKAEAKKAEAKK 51
Db 158 AKRLAAAKQAEAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAK 158
OY 52 KEVAAAEAKYAKAEAKYAKAEAKYAKAEAKYAKAEAKYAKAEAKYAKAEAKYAKAEAKYAKAEAK 85
Db 215 KAKAAAEAKAKADAEAKA--ATEAKRKADQASLD 246
RESULT 8
H1B_STRPU
ID H1B_STRPU STANDARD; PRT; 211 AA.
AC P15869;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-beta, late embryonic.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OC NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88246461; PubMed=2837660;
```

```
RA Lai Z.-C., Childs G.;
RT "Characterization of the structure and transcriptional patterns of
RT the gene encoding the late histone subtype H1-beta of the sea urchin
RT Strongylocentrotus purpuratus.";
RL Mol. Cell. Biol. 8:1842-1844 (1988).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20314; AAA30052.1; -.
DR EMBL; A28100; A28100.
DR HSSP; P02259; 1HST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
KW SEQUENCE 211 AA; 22169 MW; 9F214581334BBE7A CRC64;
Query Match 34.5%; Score 141; DB 1; Length 211;
Best Local Similarity 53.6%; Pred. No. 0.0006;
Matches 45; Conservative 5; Mismatches 30; Indels 4; Gaps 3;
OY 2 KKVAKKEKAYAKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAK 60
Db 96 KSDAQKAPDAKKAKLAKKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEK 153
OY 61 YKAEAAKAYKAEAA--KAAAEAA 83
Db 154 AKKPAKKAACKKPAKPAKAA 177
RESULT 9
MST2_DROHY
ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC Q08656;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mat101(2).
GN MST101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dmst101 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of
RT Drosophila hydei.";
RL Eur. J. Biochem. 225:1089-1095 (1994).
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
```

RESULT 12  
RS6 AEDAL





```
-----
CC EMBL; M58563; AAA27480.1; -.
DR PIR; B43592; B43592.
KW Antigen; Outer membrane; Repeat; Signal.
FT SIGNAL 21 POTENTIAL.
FT CHAIN 22 384 TREPONEMAL MEMBRANE PROTEIN B.
FT DOMAIN 151 235 17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR]-
FT [ED].
FT REPEAT 151 155 1-1.
FT REPEAT 156 160 1-2.
FT REPEAT 161 165 1-3.
FT REPEAT 166 170 1-4.
FT REPEAT 171 175 1-5.
FT REPEAT 176 180 1-6.
FT REPEAT 181 185 1-7.
FT REPEAT 186 190 1-8.
FT REPEAT 191 195 1-9.
FT REPEAT 196 200 1-10.
FT REPEAT 201 205 1-11.
FT REPEAT 206 210 1-12.
FT REPEAT 211 215 1-13.
FT REPEAT 216 220 1-14.
FT REPEAT 221 225 1-15.
FT REPEAT 226 230 1-16.
FT REPEAT 231 235 1-17.
FT DOMAIN 236 288 6 X 8 AA TANDEM REPEATS OF [EA]-A-A-R-X-
FT A-A-E.
FT REPEAT 236 243 2-1.
FT REPEAT 245 252 2-2.
FT REPEAT 254 261 2-3.
FT REPEAT 263 270 2-4.
FT REPEAT 272 279 2-5.
FT REPEAT 281 288 2-6.
SQ SEQUENCE 384 AA; 42677 MW; 6E94CB74294DE8C CRC64;

Query Match 32.6%; Score 133.5; DB 1; Length 384;
Best Local Similarity 50.0%; Pred. No. 0.0032;
Matches 44; Conservative 9; Mismatches 30; Indels 5; Gaps 3;

Oy 1 AKYAKKEKAYAKA--EKAAK--AKAYKAAEAKKAKAKAYAKAAE--KKEVA 55
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 AKKAAKKAADKAAKKAKEKAAKKAKEKAAKKAKEKAAKKAKEKAAKKAKEE 237
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 56 AAEAKYKAAEAKKAYKAAEAKKAAEAA 83
||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 AKKAAEAAKAAEAAKAAEAAKAAEAA 265
||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 15
ID -ZUO1 YEAST STANDARD; PRT; 433 AA.
AC P32527;
DT 01-OCT-1993 (Rel. 27, Created)
DD 01-OCT-1993 (Rel. 27, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zuo1in.
GN ZUO1 OR YGR285C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=20B-12;
RX MEDLINE=93010971; PubMed=1396572;
RA Zhang S., Lockshin C., Herbert A., Winter E., Rich A.;
RT "Zuo1in, a putative Z-DNA binding protein in Saccharomyces
cerevisiae";
RL EMBO J. 11:3787-3796 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97245295; PubMed=9090054;
```

```
RA Volckaert G., Voet M., Robben J.;
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT the MAL1 locus reveals 15 complete open reading frames, including
RL ZUO1, BGL2 and BIO2 genes and an ABC transporter gene.";
RL Yeast 13:251-259 (1997).
CC -!- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN
CC CHROMOSOME ORGANIZATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63612; CAA45156.1; -.
DR EMBL; Z73070; CAA97317.1; -.
DR PIR; S25194; S25194.
DR HSP; P08622; IBO2.
DR SGD; S0003517; ZUO1.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00836; DnaJ_1; 1.
DR PROSITE; PS0076; DnaJ_2; 1.
KW Chaperone; DNA-binding; Nuclear protein.
FT DOMAIN 98 170 J-DOMAIN.
FT DOMAIN 306 357 ALA/LYS-RICH.
SQ SEQUENCE 433 AA; 49019 MW; 0AA76BC11D3C7DAB CRC64;

Query Match 32.6%; Score 133.5; DB 1; Length 433;
Best Local Similarity 47.6%; Pred. No. 0.0035;
Matches 40; Conservative 12; Mismatches 25; Indels 7; Gaps 3;

Oy 2 KKYAKKEKAYAKA--EKAAK--AKAYKAAEAKKAKAKAYAKAAE--KKEVA 60
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 KEEKKEKERRKWEREAGARAEAEAKAEAEAKAEAEAKAEAEAKAEAEAKADKK 344
||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 61 YKAAEAKKAYKAAEAKKAAEAA 84
||:|||||:|||||:|||||:|||||:|||||:
Db 345 EAAKAAKKKK-K-RAIRNSAKEDY 367
||:|||||:|||||:|||||:|||||:|||||:

Search completed: March 10, 2003, 12:17:10
Job time : 19.5949 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 45.5401 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989a-6  
Perfect score: 409  
Sequence: 1 AKKYAKKEKAYAKAEKAAK.....KAYKAEAKAAKAEAYEA 86

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_rhnc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvrius:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	165.5	40.5	407	16 Q8ZQT6	Q8ZQT6 salmonella
2	163	39.9	372	2 Q9WX1	Q9WX1 pseudomonas
3	157.5	38.5	376	16 Q8Z8C1	Q8Z8C1 salmonella
4	156.5	38.3	1701	5 O61164	O61164 plasmodium
5	155	37.9	394	16 Q8X965	Q8X965 escherichia
6	154	37.7	395	2 Q937K4	Q937K4 erwinia chr
7	151	36.9	1341	16 Q98KG7	Q98KG7 rhizobium 1
8	148.5	36.3	232	10 Q9576	Q9576 chlamydomon
9	148	36.2	389	16 Q9CM70	Q9CM70 pasteurella
10	147.5	36.1	388	16 Q8ZGZ2	Q8ZGZ2 yersinia pe
11	146	35.7	1866	5 Q8T5C8	Q8T5C8 plasmodium
12	139	34.0	200	16 Q8XV7	Q8XV7 ralestonia s
13	137.5	33.6	1128	5 Q26947	Q26947 trypanosoma
14	134.5	32.9	212	3 Q93946	Q93946 candida alb
15	131.5	32.2	224	13 Q90ZD7	Q90ZD7 bufo bufo g
16	131	32.0	208	5 Q93901	Q93901 caenorhabdi

17	130.5	31.9	243	16 Q92A67	Q92A67 listeria in
18	130.5	31.9	275	5 O01395	O01395 drosophila
19	129.5	31.7	485	10 Q8RXD0	Q8RXD0 arabidopsis
20	129.5	31.7	924	10 Q9SU08	Q9SU08 arabidopsis
21	129	31.5	755	10 Q9EP71	Q9EP71 oryza sativ
22	128	31.3	1671	5 Q8T5C9	Q8T5C9 plasmodium
23	127.5	31.2	98	5 Q8WQ44	Q8WQ44 leishmania
24	127.5	31.2	111	5 Q8T9R3	Q8T9R3 leishmania
25	127.5	31.2	275	10 Q9XHL9	Q9XHL9 triticum ae
26	127.5	31.2	311	12 Q84528	Q84528 paramedum
27	127.5	31.2	619	5 Q95S18	Q95S18 drosophila
28	126.5	30.9	284	10 Q65794	Q65794 triticum ae
29	126	30.8	265	10 Q39598	Q39598 chlamydomon
30	126	30.8	288	10 Q65795	Q65795 triticum ae
31	126	30.8	845	5 Q9YIP8	Q9YIP8 plasmodium
32	125	30.6	581	16 Q9RU45	Q9RU45 deinococcus
33	124.5	30.4	909	10 Q9SU09	Q9SU09 arabidopsis
34	124.5	30.4	969	5 Q9NDI9	Q9NDI9 plasmodium
35	124	30.3	182	2 Q45370	Q45370 bordetella
36	124	30.3	298	2 Q52088	Q52088 pseudomonas
37	124	30.3	607	5 Q9V6S7	Q9V6S7 drosophila
38	123.5	30.2	191	5 Q46142	Q46142 mytilus edu
39	123	30.1	356	16 Q9XR10	Q9XR10 vibrio chol
40	122.5	30.0	191	5 Q46141	Q46141 mytilus edu
41	122.5	30.0	191	5 Q46362	Q46362 mytilus edu
42	122.5	30.0	237	10 Q9SWU2	Q9SWU2 triticum ae
43	122.5	30.0	295	10 Q9ZR20	Q9ZR20 pisum sativ
44	122.5	30.0	295	10 Q9AT22	Q9AT22 lathyrus sa
45	122.5	30.0	297	10 Q9SXQ8	Q9SXQ8 pisum sativ

ALIGNMENTS

RESULT 1

Q8ZQT6	ID	Q8ZQT6	PRELIMINARY;	PRT;	407 AA.
AC	Q8ZQT6	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)				
DE	Toi protein, membrane spanning protein.				
GN	TOLA OR STM0747.				
OS	Salmonella typhimurium.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_TaxID=602;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,				
RA	Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2."				
RL	Nature 413:852-856(2001).				
DR	EMBL; AE008730; AAL19691.1; -				
DR	InterPro; IPR00104; Antifreeze_1.				
DR	PRINTS; PR00308; ANTIFREEZE1.				
KW	Complete proteome.				
SQ	SEQUENCE 407 AA; 41865 MW; 2534352116602D75 CRC64;				

Query Match 40.5%; Score 165.5; DB 16; Length 407;  
Best Local Similarity 51.9%; Pred. No. 2.5e-05;  
Matches 55; Conservative 14; Mismatches 16; Indels 21; Gaps 6;

Qy 1 AKKYAKKEKAYAKAEKAA-----KKAAYK-AAEAKKKAKAEKKA-----KAAKAEK 51

Db 123 AAKLAQQQQQAEAAKAAADAKKKAEAAKAAADAKKKAEAAVAAADAKKKAEAA 182

Qy 52 KEYAAAEAKYKAE-----AKKAYKAAEAKAAA---KEAAAEA 86  
: |||: ||| ||| : ||||| ||| |: | ||  
Db 183 AK-AAADAKKKKAEAEAAKAAAEAKKKAEAEAAKAAAEAKKKADAA 227

```

RESULT 2
ID Q9WWX1 PRELIMINARY; PRT; 372 AA.
AC Q9WWX1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Tola protein.
DE TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RN [2]
RX MEDLINE=96198174; PubMed=8626299;
RT Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
RT lipoprotein (PAL) is involved in maintenance of the integrity of the
RT cell envelope.";
RT J. Bacteriol. 178:1699-1706(1996).
RN [3]
RP SEQUENCE FROM N.A.
RN [4]
RX MEDLINE=96422022; PubMed=8824639;
RT Rodriguez-Herva J.J., Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida.";
RT J. Bacteriol. 178:5836-5840(1996).
RN [5]
RX EMBL; X74218; CAB50780.1; -.
RN [6]
RT Interpro; IPR001386; Histone_H1/H5.
RN [7]
RX PRINTS; PRO0624; HISTONEH5.
RN [8]
RX SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64

```

Qy		50 EKKEYAAAEAKYK-AEAAKKAYKAEEAATAAKEAAEYA 86    :    :    :      - :     :    :   :
Db		208 AKKAAAEDAKKAAEAEEKKAAAAOAOKKQAQEARKA 245

RESULT 3	Q828C1	PRELIMINARY;	PRT;	376 AA.
ID	Q828C1			
AC	Q828C1			
DT	01-MAR-2002 (TReMBLrel. 20, Created).			
DT	01-MAR-2002 (TReMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	TolA protein.			
GN	STX0793			
OS	<i>Salmonella typhi</i> .			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			

[illegible]

```

RESULT 4
ID 061164
ID 061164 PRELIMINARY; PRT; 1701 AA.
AC 061164;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Erythrocyte binding protein.
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YN;
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria
RT parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235 (1998).
DR ENBL; AF031886; AAC05366.1; -.
SQ SEQUENCE 1701 AA; 139268 MW; EDABE2BDFD87CE8AFCRC64;

```

Query Match	38.3%	Score 156.5	DB 5	Length 1701
Best Local Similarity	53.5%	Pred. No. 0.00048		
Matches 53	Conservative 9	Mismatches 18	Indels 19	Gaps 6

```

RESULT 5
Q8X965
ID Q8X965 PRELIMINARY; PRT; 394 AA.
AC Q8X965;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Membrane spanning protein, required for outer membrane integrity
DE (Membrane spanning protein TolA).
DE GN TOLA OR Z0907 OR EGS0774.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=83334;
RX [1]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RX [2]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa G., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005252; AAG55075.1; -.
DR ENBL; AP002553; BAB34197.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 40517 MW; 5B58D8E8230BDE28 CRC64;

Query Match 37.9%; Score 155; DB 16; Length 394;
Best Local Similarity 53.1%; Pred. No. 0.00015;
Matches 53; Conservative 6; Mismatches 23; Indels 16; Gaps

Qy 1 AKKYAKKEKAYA-----KKAEKAA-----KKAEKAYKAAEAKKKAAEA-----KKYAKAA 47
Db 167 AKKAAEAAKAAEAQKKAEEAAALKKKAAEAEEAAEAARKKAAEAADKKAAEA 226

Qy 48 KAEKKEVAAAEKAYKAAEKYKAAEAAKAAAEKAA 83
Db 227 AAEEK---AAADKKAAEAADKKKAAEAAKAAEA 259

RESULT 6
Q937K4
ID Q937K4 PRELIMINARY; PRT; 395 AA.
AC Q937K4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Tola protein.
DE GN TOLA.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OC NCBI_TaxID=556;
RX [1]
SEQUENCE FROM N.A.
RC STRAIN=3937;

```









GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 44.8143 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-6  
Perfect score: 409  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....KKAYKAEAKAAKAEAYEA 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	409	100.0	86	21 AAY82576	Copolymer molecule
2	351.5	85.9	77	21 AAY82575	Copolymer molecule
3	312.5	76.4	109	21 AAY82577	Copolymer molecule
4	268	65.5	66	21 AAY82574	Copolymer molecule
5	233	57.0	56	21 AAY82573	Copolymer molecule
6	163	39.9	154	11 AAR06445	Recombinant copoly
7	157	38.4	106	11 AAR06446	Recombinant copoly
8	156	38.1	100	21 AAY98499	Peptide #10 used i
9	156	38.1	100	21 AAY59044	Amino acid polymer
10	156	38.1	100	22 AAU04289	Poly-Lys-Ala used

11	156	38.1	100	22 AAB45852	Nucleic acid trans
12	147	35.9	214	20 AAY34055	M. tuberculosis hi
13	147	35.9	214	21 AAY57353	M. tuberculosis hi
14	146.5	35.8	205	21 AAB20575	Mycobacterium bovi
15	137.5	33.6	472	16 AAB84569	Trypanosoma cruzi
16	137.5	33.6	564	16 AAR84565	Trypanosoma cruzi
17	137.5	33.6	643	16 AAR84568	Trypanosoma cruzi
18	135	33.0	140	13 AAR25206	Synthetic helical
19	133.5	32.6	433	18 AAW30256	Zuotin. Saccharom
20	133.5	32.6	433	20 AAY32954	Human zuotin prote
21	133.5	32.6	433	22 AAG70739	S cerevisiae apopt
22	129.5	31.7	223	20 AAY14928	Amino acid sequenc
23	129	31.5	219	21 AAY76981	Quail Hi histone p
24	127.5	31.2	1507	21 AAB24128	Plasmodium yoelii
25	127	31.1	222	18 AAW29476	Human histone H1 i
26	127	31.1	222	20 AAY34033	Human histone H1 i
27	127	31.1	222	21 AAY57331	Human histone H1 i
28	126.5	30.9	45	21 AAY82572	Copolymer molecula
29	124	30.3	334	22 ABG28693	Novel human diagno
30	124	30.3	607	22 ABB62173	Drosophila melanog
31	123	30.1	130	23 AAU09945	Peptide fragment o
32	123	30.1	130	23 AAE13237	Human linker histo
33	121	29.6	158	20 AAY34058	Histone H1 isoform
34	121	29.6	158	21 AAY57366	Human histone H1.5
35	121	29.6	226	20 AAY34060	Human histone H1 i
36	121	29.6	226	21 AAY57358	Human histone H1 i
37	121	29.6	239	23 ABB49123	Listeria monocytog
38	119.5	29.2	332	21 AAY75540	Neisseria meningit
39	119.5	29.2	332	21 AAY75541	Neisseria meningit
40	119	29.1	234	23 AAU09944	Histone H1-4, frag
41	119	29.1	234	23 AAE13234	Human linker histo
42	118.5	29.0	111	18 AAW44829	Adhesive polypepti
43	117	28.6	218	18 AAW29477	Human histone H1 i
44	117	28.6	218	20 AAY34034	Human histone H1 i
45	117	28.6	218	21 AAY57332	Human histone H1 i

ALIGNMENTS

RESULT 1  
AAY82576  
ID AAY82576 standard; peptide; 86 AA.  
XX AAY82576;  
AC AAY82576;  
DT 28-JUL-2000 (first entry)  
XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.  
XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX Unidentified.  
XX OS  
XX WO200018794-A1.  
XX PD  
XX 06-APR-2000.  
XX PF  
XX 24-SEP-1999; 99WO-US22402.  
XX PR  
XX 25-SEP-1998; 98US-0101693.  
XX PA  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX (TEVA-) TEVA PHARM USA INC.



inflammatory condition; multiple sclerosis; rheumatoid arthritis; Crohn's disease; chronic immune thrombocytopenia purpura; colitis; diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis; Hashimoto's disease; idiopathic myxoedema; myasthenia gravis; pemphigus vulgaris; systemic lupus erythematosus.

Unidentified.

WO200018794-A1.

06-APR-2000.

24-SEP-1999; 99WO-US22402.

25-SEP-1998; 98US-0101693.

(YEDA ) YEDA RES & DEV CO LTD.  
(TEVA-) TEVA PHARM USA INC.

Gad A, Lis D;

WPI; 2000-317499/27.

Copolymer 1 related polypeptides used as molecular weight markers for glatiramer acetate and for treatment and prevention of immune diseases

Claim 10; Page 14; 72pp; English.

AAV82571 to AAY82577 represent specifically claimed copolymer molecular weight TV-marker polypeptides from the present invention. The present invention describes polypeptides (I) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glatiramer acetate related tetrapolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, demyelinating diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's disease, chronic immune thrombocytopenia purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical properties which are analogous to glatiramer acetate molecules, which makes them ideal for use as molecular weight markers.

Sequence 109 AA;

Query Match 76.4%; Score 312.5; DB 21; Length 109;  
Best Local Similarity 72.1%; Pred. NO. 2e-21;  
Matches 80; Conservative 1; Mismatches 3; Indels 27; Gaps 5;

QY 1 AKKYAKK-EKAYAKA-----EKAKKAEKAYKAAEKKA----- 36  
DB 1 AKKYAKKAEKAYKAAEKKA EKAYKAEKAYKAAEKKA EKAYKAEKAYKAAEKKA 60

QY 37 -KAEAKKYAKAAKKEKYAAAEKAYKAAEKKA EKAYKAAEKKA EKAYKAAEKKA 86  
DB 61 YKAEAKKYAKAAKKEKYAAAEK-KAEAA-KAYKAAEKKA EKAYKAAEKKA 109

RESULT 4  
ID AAY82574  
XX AAY82574 standard; peptide; 66 AA.  
AC AAY82574;  
XX

28-JUL-2000 (first entry)

Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.

Copolymer; molecular weight marker; TV-marker; immune disease; glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective; osteopathic; immunosuppressive; antithyroid; antiinflammatory; antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological; antinaeamic; immunosuppressive; demyelinating disease; osteoarthritis; inflammatory condition; multiple sclerosis; rheumatoid arthritis; Crohn's disease; chronic immune thrombocytopenia purpura; colitis; diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis; Hashimoto's disease; idiopathic myxoedema; myasthenia gravis; pemphigus vulgaris; systemic lupus erythematosus.

Unidentified.

WO200018794-A1.

06-APR-2000.

24-SEP-1999; 99WO-US22402.

25-SEP-1998; 98US-0101693.

(YEDA ) YEDA RES & DEV CO LTD.  
(TEVA-) TEVA PHARM USA INC.

Gad A, Lis D;

WPI; 2000-317499/27.

Copolymer 1 related polypeptides used as molecular weight markers for glatiramer acetate and for treatment and prevention of immune diseases

Claim 10; Page 14; 72pp; English.

AAV82571 to AAY82577 represent specifically claimed copolymer molecular weight TV-marker polypeptides from the present invention. The present invention describes polypeptides (I) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glatiramer acetate related tetrapolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, demyelinating diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's disease, chronic immune thrombocytopenia purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical properties which are analogous to glatiramer acetate molecules, which makes them ideal for use as molecular weight markers.

Sequence 66 AA;

Query Match 65.5%; Score 268; DB 21; Length 66;  
Best Local Similarity 75.6%; Pred. No. 1.3e-17;  
Matches 65; Conservative 0; Mismatches 1; Indels 20; Gaps 3;

QY 1 AKKYAKKAEKAYKAAEKKA EKAYKAAEKKA EKAYKAAEKKA EKAYKAAEKKA 60  
DB 1 AKKYAKKAEKAYK-----AKKAEK-----AAKKAEKKA EKAYKAAEKKA 49

QY 61 YKAAKAYKAAEKKA EKAYKAAEKKA EKAYKAAEKKA EKAYKAAEKKA 86





```
Best Local Similarity 55.3%; Pred. No. 2.7e-07;
Matches 47; Conservative 6; Mismatches 30; Indels 2; Gaps 2;

QY 1 AKKYAK-KEKAYAKAEKAAKAKAKAYKAEAKKAKAKAEAKKYAKA-AKAEKKEVAAAE 58
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61

QY 59 AKYKAEAAKAYKAEAAKAAAEAA 83
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 86

RESULT 9
AAV59044
ID AAV59044 standard; peptide; 100 AA.
XX
AC AAV59044;
XX
DT 07-MAR-2000 (first entry)
XX
DE Amino acid polymer seq ID NO: 64 of US5994109.
XX
KW Nucleic acid transport system; NTS; cell surface receptor; cytosol;
KW nuclear membrane; lysis moiety; transgenic animal; human disease;
KW nucleic acid delivery; cancer.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 3..100
FT /note= "Lys-Ala in positions 3 to 100 may be optionally
FT absent"
XX
XX US5994109-A.
XX
XX 30-NOV-1999.
XX
XX 03-JUN-1995; 95US-0460890.
XX
XX 14-DEC-1993; 93US-0167641.
XX
XX 20-MAR-1992; 92US-0855389.
XX
XX 19-MAR-1993; 93WO-US02725.
XX
XX 14-DEC-1993; 93US-0167641.
XX
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX Woo SLC, Cristiano RJ, Gottchalk S, Sparrow J, Smith LC;
XX WPI; 2000-038262/03.
XX
XX Nucleic acid transport system, useful for creating transgenic animals
XX for assessing human disease such as cancer in an animal model -
XX
XX Disclosure; Columns 123-124; 107pp; English.
XX
XX The invention relates to a nucleic acid transport system (NTS) for
XX delivering nucleic acid into a cell. The NTS contains but is not limited
XX to 5 components: (a) the nucleic acid or a macromolecule to be delivered;
XX (b) a moiety that recognizes and binds to a cell surface receptor or
XX antigen or is capable of entering a cell through cytosol; (c) a nucleic
XX acid or macromolecular molecule binding moiety; (d) a moiety that is
XX capable of moving or initiating movement through a nuclear membrane; and/
XX or (e) a lysis moiety that enables the transport of the entire complex
XX from the cell surface directly into the cytoplasm of the cell. The NTS
XX delivers nucleic acid into the cellular interior as well as the nucleus
XX of specific cells. The NTS can be used to treat disorders by targeting
XX specific nucleic acid accordingly. The NTS can also be used to create
XX transgenic animals for assessing human disease, such as cancer, in an
XX animal model. The NTS can be used in vitro with tissue culture cells
XX which allows the role of various nucleic acids to be studied by targeting
XX specific expression into specifically targeted tissue culture cells. The
XX lysis agent within the NTS avoids the problem of endosomal/lysosomal
XX degradation.
```

```
XX
SQ Sequence 100 AA;
Query Match 38.1%; Score 156; DB 21; Length 100;
Best Local Similarity 55.3%; Pred. No. 2.7e-07;
Matches 47; Conservative 6; Mismatches 30; Indels 2; Gaps 2;

QY 1 AKKYAK-KEKAYAKAEKAAKAEAKKAYKAAAEAKKAKAKAEAKKYAKA-AKAEKKEVAAAE 58
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 61

QY 59 AKYKAEAAKAYKAEAAKAAAEAA 83
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 86

RESULT 10
AAU04289
ID AAU04289 standard; Peptide; 100 AA.
XX
AC AAU04289;
XX
DT 23-OCT-2001 (first entry)
XX
DE Poly-Lys-Ala used in nucleic acid transporter system.
XX
KW Nucleic acid transport; cytosol; ligand; lysis agent; spacer molecule;
KW gene therapy; hepatocyte; muscle; bone forming cell.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 3..100
FT /note= "Lys-Ala in positions 3-100 may be present or
FT absent"
XX
XX US6177554-B1.
XX
XX 23-JAN-2001.
XX
XX 05-JUN-1995; 95US-0462040.
XX
XX 14-DEC-1993; 93US-0167641.
XX
XX 20-MAR-1992; 92US-0855389.
XX
XX 19-MAR-1993; 93WO-US02725.
XX
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX Woo SLC, Smith LC, Cristiano RJ, Gottchalk S, Sparrow J;
XX WPI; 2001-365933/38.
XX
XX Nucleic acid transport system, useful for creating transgenic animals
XX for assessing human disease such as cancer in an animal model -
XX
XX Disclosure; Column 131; 111pp; English.
XX
XX The sequence represents poly-Lys-Ala, used to bind nucleic acid in a
XX nucleic acid transporter system. The nucleic acid transporter system uses
XX nucleic acid binding complexes containing surface ligands which are
XX capable of binding to a cell surface receptor and entering the cell
XX through cytosol. The compounds of the invention are either ligands,
XX binding molecules (surface ligands), lysis agents, spacer molecules or
XX their intermediates. The ligands, binding molecules, lysis agents and
XX spacer molecules are used in nucleic acid transporter systems to deliver
XX nucleic acid into specific cells e.g. in gene therapy to deliver nucleic
XX acid into hepatocytes, muscle cells or bone forming cells.
XX
XX Sequence 100 AA;
Query Match 38.1%; Score 156; DB 22; Length 100;
Best Local Similarity 55.3%; Pred. No. 2.7e-07;
Matches 47; Conservative 6; Mismatches 30; Indels 2; Gaps 2;
```







```
XX 20-MAR-1995; 95WO-US03191.
XX
XX 24-MAR-1994; 94US-0216894.
XX
XX (KIRC/) KIRCHHOFF L V.
XX (OTSU/) OTSU K.
XX
XX Kirchhoff LV, Otsu K;
XX
XX WPI; 1995-344618/44.
XX N-PSDB; AAT05333.
XX
XX New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as
XX immunassay reagent for specific diagnosis of Chagas disease, also
XX related nucleic acid and transformed cells
XX
XX Disclosure; Page 44-45; 68pp; English.
XX
XX AAR84565-R84569 are polypeptides of the TCR27 protein of T.cruzi
XX The proteins are all fusion products with glutathione-S-transferase
XX (GST) and some contain a linker sequence. The TCR27 protein comprises
XX a 95 amino acid (aa) N-terminal region; 69 repeats of a highly
XX conserved 14 aa sequence and a 68 aa C-terminal region. This sequence
XX encodes the GST sequence, the Ag8 polypeptide contg. 15 of the 69
XX repeat units and also contains a linker sequence. The TCR27
XX polypeptides of the invention are useful for the diagnosis of Chagas
XX disease (American Trypanosomiasis), they are capable of detecting
XX anti-T.cruzi antibodies; or for blood screening. The TCR27 protein
XX has epitopes to which most T.cruzi infected individuals have
XX antibodies. The TCR27 polypeptides will not react with serum from
XX patients with leishmaniasis, schistosomiasis, or autoimmune disease
XX and are hence less likely to cause false positives in diagnosis.
XX
XX Sequence : 472 AA;

Query Match 33.6%; Score 137.5; DB 16; Length 472;
Best Local Similarity 48.9%; Pred. No. 6.9e-05;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

Qy 1 AKKYAKKEKAYAKKAEAKKAEAKYKAAEAKKAKAEAKKYAKAAK-AEKKEYAAAEA 59
Db 346 ATKVAEEK--OKAAEATKVAAEKQKAAEATKVAAEKQKAAEATKVAAEKQKAAEA 402
Qy 60 KYKAAAK-KAYKAAKAAKAAKAAVEA 86
Db 403 TKVAAEKQKAAEATKVAAEKQKAAEA 430
```

Search completed: March 10, 2003, 12:21:30  
Job time : 44.8143 secs

THIS PAGE BLANK (USPTO)











Search completed: March 10, 2003, 12:53:44  
Job time : 25.7251 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 15.2405 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-6

Perfect score: 409

Sequence: 1 AKYAKKEKAYAKAEKAAK.....KKAYKAEAKAAKAEAYEA 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	156	38.1	100	2	US-08-460-890A-64
2	156	38.1	100	3	US-08-167-641C-64
3	156	38.1	100	4	US-08-460-971A-64
4	156	38.1	100	4	US-08-462-040-64
5	147	35.9	214	3	US-09-041-889-27
6	137.5	33.6	472	2	US-08-216-894-10
7	137.5	33.6	472	4	US-09-115-746-10
8	137.5	33.6	564	2	US-08-216-894-2
9	137.5	33.6	564	4	US-09-115-746-2
10	137.5	33.6	643	2	US-08-216-894-8
11	137.5	33.6	643	4	US-09-115-746-8
12	133.5	32.6	433	1	US-08-346-849-2
13	133.5	32.6	433	2	US-08-293-284A-2
14	129.5	31.7	223	4	US-09-095-855-201
15	129.5	31.7	223	4	US-09-205-426-201
16	127.5	31.2	1507	3	US-08-929-329-5
17	127	31.1	222	3	US-09-041-889-3
18	127	31.1	222	3	US-08-837-058-3
19	121	29.6	158	3	US-09-041-889-40
20	121	29.6	226	3	US-09-041-889-32
21	119.5	29.2	56	4	US-08-993-008A-6
22	117	28.6	218	3	US-09-041-889-4
23	117	28.6	218	3	US-08-837-058-4
24	115.5	28.2	262	1	US-08-403-379A-1
25	115.5	28.2	262	2	US-08-929-414-1
26	115.5	28.2	263	2	US-08-557-309B-51
27	115.5	28.2	263	3	US-08-834-306-51

28	115.5	28.2	263	4	US-08-993-674A-51	Sequence 51, Appl
29	115.5	28.2	263	4	US-09-256-976-51	Sequence 51, Appl
30	109	26.7	100	2	US-08-460-890A-62	Sequence 62, Appl
31	109	26.7	100	3	US-08-167-641C-62	Sequence 62, Appl
32	109	26.7	100	4	US-08-460-971A-62	Sequence 62, Appl
33	109	26.7	100	4	US-08-462-040-62	Sequence 62, Appl
34	109	26.7	434	1	US-08-097-830B-3	Sequence 3, Appl
35	109	26.7	434	2	US-08-456-112B-3	Sequence 3, Appl
36	106.5	26.0	48	4	US-08-993-008A-5	Sequence 5, Appl
37	105.5	25.8	219	2	US-08-557-309B-54	Sequence 54, Appl
38	105.5	25.8	442	3	US-08-834-306-52	Sequence 52, Appl
39	105.5	25.8	442	4	US-08-993-674A-52	Sequence 52, Appl
40	105.5	25.8	442	4	US-09-256-976-52	Sequence 52, Appl
41	105	25.7	92	4	US-09-344-529-2	Sequence 2, Appl
42	105	25.7	1507	6	US-09-256-976-52	Sequence 2, Appl
43	104	25.4	288	3	US-08-312-949-4	Sequence 4, Appl
44	104	25.4	288	3	US-08-446-201-4	Sequence 4, Appl
45	104	25.4	289	1	US-08-072-070-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-460-890A-64  
; Sequence 64, Application US/08460890A  
; Patent No. 5994109  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066

COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,890A  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/167,641  
; FILING DATE: December 14, 1993  
; APPLICATION NUMBER: 07/855,389  
; FILING DATE: March 20, 1992  
; APPLICATION NUMBER: PCT/US93/02725  
; FILING DATE: March 19, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 212/066  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single





```

1  APPLICANT: Otsu, Keiko
2  TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
3  TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
4  NUMBER OF SEQUENCES: 10
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Foley & Lardner
7  STREET: 3000 K Street, N.W., Suite 500
8  CITY: Washington, D.C.
9  COUNTRY: USA
10 ZIP: 20007-5109
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.25
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/216,894
18 FILING DATE: 24-MAR-1994
19 ATTORNEY/AGENT INFORMATION:
20 NAME: BENT, Stephen A.
21 REGISTRATION NUMBER: 29,768
22 REFERENCE/DOCKET NUMBER: 85326/102/DRLO
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (202)672-5300
25 TELEFAX: (202)672-5399
26 TELEX: 904136
27 INFORMATION FOR SEQ ID NO: 10:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 472 amino acids
30 TYPE: amino acid
31 TOPOLOGY: linear
32 MOLECULE TYPE: protein
33
34 US-08-216-894-10

```

```

Query Match          33.6%; Score 137.5; DB 2; Length 472;
Best Local Similarity 48.9%; Pred.No. 9.5e-06;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

QY      1 AKKYAKERKAYAKKAEEAAKAEKAYKAAEAKKKAKAEEAKKVAKAAK--AEKEYAAAAEA 59
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       346 ATKVAEAEK--QKAEATKVAAEAKKQKAAEATKVAAEKQKAAEATKVAAEKQKAAEA 402
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY      60 KYKAAEAK-KAYKAAEAKAAKAAKEAAAEA 86
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       403 TKVAAEAKQKAAEATKVAAEAKQKAAEA 430
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 7
US-09-115-746-10
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Oesu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-10

Query Match          33.6%; Score 137.5; DB 4; Length 472;
Best Local Similarity 48.9%; Pred. No. 9.5e-06;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps
                                5;

Qy      1 AKKYAKKEKAVAKAEAKAAKKAAKAVKAAEAKKKAKAEATKVAAEKQKAEEA
Db      346 ATKVAAEK--QKAEBATKVAAEKQKAEEATKVAAEKQKAEEATKVAAEKQKAEEA
Qy      60 KYKAAEA-KAYKAAEAKAAKAAKEAAVEA 86
Db      403 TKVAAEKQKAEEATKVAAEKQKAEEA 430

RESULT 8
US-08-216-894-2
; Sequence 2, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-2

Query Match          33.6%; Score 137.5; DB 2; Length 564;
Best Local Similarity 48.9%; Pred. No. 1.1e-05;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps
                                5;

```



```
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-8

Query Match 33.6%; Score 137.5; DB 4; Length 643;
Best Local Similarity 48.9%; Pred. No. 1.3e-05;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

Qy 1 AKKYAKKEKAYAKKAAKAAKAYKAAAEAKKKAKAEAKKYAKAAK-AEKKEYAAAE 59
Db 462 ATKVAEEK---QAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 518

Qy 60 KYKAAAK-KAYKAAAKAAKAAKAAEA 86
Db 519 TKVAEAEKQKAAEATKVAEAEKQKAAEA 546

RESULT 12
US-08-346-849-2
; Sequence 2, Application US/08346849
; Patent No. 5670483
; GENERAL INFORMATION:
; APPLICANT: Zhang, Shuguang
; APPLICANT: Lockshin, Curtis
; APPLICANT: Rich, Alexander
; APPLICANT: Holmes, Todd
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,849
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28 DECEMBER 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 32.6%; Score 133.5; DB 2; Length 433;
Best Local Similarity 47.6%; Pred. No. 2e-05;
Matches 40; Conservative 12; Mismatches 25; Indels 7; Gaps 3;

Qy 2 KKYAKKEKAYAKKAAKAAKAYKAAAEAKKKAKAEAKKYAKA-AKAEKKEYAAAEAK 60
Db 290 KEEKKEKERRKWEREAGARAEAEAKAKAEAEAKAEAEAKAKASAKADKK-----KAK 344

Qy 61 YKAEAAKKAYKAAKAAKAAKAAEA 84
Db 345 EAAKAAKKK-NK-RAIRNSAKEADY 367

; MOLECULE TYPE: protein
US-08-346-849-2

Query Match 32.6%; Score 133.5; DB 1; Length 433;
Best Local Similarity 47.6%; Pred. No. 2e-05;
Matches 40; Conservative 12; Mismatches 25; Indels 7; Gaps 3;

Qy 2 KKYAKKEKAYAKKAAKAAKAYKAAAEAKKKAKAEAKKYAKA-AKAEKKEYAAAEAK 60
Db 290 KEEKKEKERRKWEREAGARAEAEAKAKAEAEAKAEAEAKAKASAKADKK-----KAK 344

Qy 61 YKAEAAKKAYKAAKAAKAAKAAEA 84
Db 345 EAAKAAKKK-NK-RAIRNSAKEADY 367
```

Db 345 EAAGAAGGKKNK-RAIRNSAKEADY 367

## RESULT 14

US-09-855-201  
; Sequence 201, Application US/09095855  
; Patent No. 6160093

GENERAL INFORMATION:  
; APPLICANT: Tan, Paul

APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot

APPLICANT: Prestidge, Ross  
; TITLE OF INVENTION: Compounds and Methods for

TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections  
; NUMBER OF SEQUENCES: 208

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle

STATE: WA  
COUNTRY: USA

ZIP: 98121  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855

FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347

FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970

FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362

FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c3  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563

TELEX:  
INFORMATION FOR SEQ ID NO: 201:

SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-855-201

Query Match 31.7%; Score 129.5; DB 4; Length 223;

Best Local Similarity 54.3%; Pred. No. 2.4e-05;  
Matches 50; Conservative 5; Mismatches 26; Indels 11; Gaps 6;

Qy 1 AKKYAKK---EKAYAKKAEKAAKAEAK--AYKAAEAKKKAKAEAKKYAKAAKAEKKEYA 55

Db 112 ARKAAPAKPAKAAKAAKKA-APAKAPAKKAATKAAPAKKATAAKKAAAPAKKATAAKK--- 167

Qy 56 AAEAKYKAEAAKAYK-AEAAKAAKAAEAAEAA 86

Db 168 AAPAK-KAPAKKAATKAAPAKKAPAKKATKA 198

## RESULT 15

US-09-426-201

; Sequence 201, Application US/09205426  
; Patent No. 6406704

GENERAL INFORMATION:  
; APPLICANT: Watson, James D.

APPLICANT: Tan, Paul L. J.  
; TITLE OF INVENTION: Compounds and Methods for Treatment and

TITLE OF INVENTION: Diagnosis of Mycobacterial Infections  
; FILE REFERENCE: 11000.1002c4

CURRENT APPLICATION NUMBER: US/09/205,426  
CURRENT FILING DATE: 1998-12-04

EARLIER APPLICATION NUMBER: 09/095,855  
EARLIER FILING DATE: 1998-06-11

EARLIER APPLICATION NUMBER: 08/997,362  
EARLIER FILING DATE: 1997-12-23

EARLIER APPLICATION NUMBER: 08/873,970  
EARLIER FILING DATE: 1997-06-12

EARLIER APPLICATION NUMBER: 08/705,347  
EARLIER FILING DATE: 1996-08-29

NUMBER OF SEQ ID NOS: 208  
SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 201  
LENGTH: 223

TYPE: PRT  
ORGANISM: Mycobacterium vaccae

US-09-205-426-201

Query Match 31.7%; Score 129.5; DB 4; Length 223;

Best Local Similarity 54.3%; Pred. No. 2.4e-05;  
Matches 50; Conservative 5; Mismatches 26; Indels 11; Gaps 6;

Qy 1 AKKYAKK---EKAYAKKAEKAAKAEAK--AYKAAEAKKKAKAEAKKYAKAAKAEKKEYA 55

Db 112 ARKAAPAKPAKAAKAAKKA-APAKAPAKKAATKAAPAKKATAAKKATAAKKATAAKK--- 167

Qy 56 AAEAKYKAEAAKAYK-AEAAKAAKAAEAAEAA 86

Db 168 AAPAK-KAPAKKAATKAAPAKKAPAKKATKA 198

Search completed: March 10, 2003, 12:30:03

Job time : 15.2405 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 31.9641 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989A-7  
Perfect score: 519  
Sequence: 1 AKKYAKKAEKAYAKKAAK.....AKAYKAEAKAAKAAKAAAYEA 109

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	189.5	36.5	421	JV0057	tolA protein - Eac
2	181.5	35.0	394	F90725	membrane spanning
3	181.5	35.0	394	G85576	membrane spanning
4	181	34.9	210	A25550	histone H1 - sea u
5	179	34.5	347	E83525	tolA protein PA097
6	178	34.3	376	AC0592	tolA protein [impo
7	178	34.3	388	AC0138	tolA colicin impor
8	175.5	33.8	344	S34153	mst101-1 protein -
9	172.5	33.2	211	A28100	histone H1-beta, e
10	171.5	33.0	1390	S51364	sperm tail-specific
11	170.5	32.9	248	HSUR1P	histone H1, gonada
12	170.5	32.9	1701	T09127	probable erythrocy
13	165	31.8	220	A28456	histone H1.10 - ch
14	164.5	31.7	231	S59589	histone H1 - Chlam
15	164	31.6	265	S19113	cgsr-4 protein - C
16	164	31.6	311	T17698	hypothetical prote
17	163.5	31.5	384	B43592	outer membrane pro
18	163	31.4	182	S61926	histone H1 homolog
19	163	31.4	225	B28456	histone H1.11L - c
20	162.5	31.3	1128	T30296	R27-2 protein - Tr
21	162	31.2	356	A82152	tolA protein VC183
22	161.5	31.1	328	A44993	cytosolic repetiti
23	161	31.0	219	E60110	repetitive protein
24	160.5	30.9	206	HSR1R	histone H1 - rainb
25	160.5	30.9	226	S51660	histone H1-5 [vali
26	159.5	30.7	219	C28456	histone H1.11R - c
27	158	30.4	219	HSU1B	histone H1-4 [vali
28	157	30.3	771	A33430	h-caldesmon - chic
29	156.5	30.2	219	A23055	histone H1.01 - ch

30	154.5	29.8	217	2	S29309	hypothetical prote
31	154.5	29.8	241	2	JN0748	histone H1-II - Vo
32	154.5	29.8	309	2	G83013	polyhydroxyalkanoa
33	153.5	29.6	369	2	H71321	conserved hypothet
34	153.5	29.6	372	2	G64064	outer membrane int
35	152.5	29.4	224	2	D28456	histone H1.03 - Ch
36	151	29.1	1403	2	T11583	probable translati
37	150	28.9	212	2	A28470	histone H1 - mouse
38	150	28.9	217	2	A26721	histone H1 -gamma,
39	150	28.9	288	2	T06257	histone H1 (clone
40	148	28.5	218	2	S01262	histone H1 - musco
41	147.5	28.4	219	2	I49742	histone H1 - mouse
42	147	28.3	208	2	T23778	histone H1.1 - Cae
43	147	28.3	221	2	S49482	histone H1 - mouse
44	147	28.3	243	2	AE1689	hypothetical prote
45	147	28.3	340	2	A35630	regulatory protein

ALIGNMENTS

RESULT 1

JV0057  
tolA protein - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002  
C:Accession: JV0057; B64810  
R:Levengood, S.K.; Webster, R.R.  
J: Bacteriol. 171, 6600-6609, 1989  
A:Title: Nucleotide sequences of the tolA and tolB genes and localization of their produ  
A:Reference number: JV0057; PMID:90078104; PMID:2687247  
A:Accession: JV0057  
A:Molecule type: DNA  
A:Residues: 1-421 <LEV>  
A:Cross-references: GB:M28232; NID:g148018; PIDN:AAA24683.1; PID:g148019  
A:Experimental source: strain JM105  
A:Note: The authors translated the initiation codon GTG for residue 1 as Val  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; PMID:97428617; PMID:9278503  
A:Accession: B64810  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-421 <BLAT>  
A:Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AACT3833.1; PID:g1786960;  
A:Experimental source: strain K-12, substrain MG1655  
C:Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach t  
C:Genetics:  
A:Gene: tolA  
A:Map position: 17 min  
A:Start codon: GTG  
A:Keywords: nucleotide binding; P-loop; transmembrane protein  
F:14-34/Domain: transmembrane #status predicted <MSS>  
F:78-301/Domain: helical #status predicted <HSR>  
F:355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 36.5%; Score 189.5; DB 2; Length 421;  
Best Local Similarity 53.0%; Pred. No. 1.5e-05;  
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5;

QY 1 AKKYAKKAE---KAYAKKAAKKEKAYAKKAAKKAAYKAAKKAAYAK-EEAK 55  
Db 120 AEEAKQAEELKQKQAEAAKAAKAAADAKAKAADAKA---AEEAKKAAADAKKKAEEAAK 177  
QY 56 AKKEAY-KAEAKKYAKAAKKEKKEAYAAAEAKK---AEAKAYKAAKAAKAAKAAK 106  
Db 178 AAEEAQKKAEEAAAKKKKAAEAAAEAAAEAKKAATEAAEAKAEKKAEEAKAAEKA 232

RESULT 2

F90725

membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
A:Accession: F90725  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A:Reference number: A99629; UID:21156231; PMID:11258796  
A:Accession: F90725  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-394 <NA>  
A:Cross-references: GB:BA000007; PIDN:BA034197.1; PID:G13360233; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 050952  
C:Genetics:  
A:Gene: EC0774

Query Match 35.0%; Score 181.5; DB 2; Length 394;  
Best Local Similarity 52.2%; Pred. No. 4.7e-05;  
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

QY 1 AKYAKKAE----KAYAKKAAKAEKKAYAKKEAKYAKAEAKKKAKAEAKKYAK-BAAK 55  
DB 120 AEEAAQAEELKQKQAEAAKAAADAKAKAEADDKA--AEAAKAAADAKKAEAEAAK 177  
QY 56 AKKEAY-KAEAKKYAKAKAEKKEVAAAEAKK---AEAAKAYKAEAAKAAKEAA 106  
DB 178 AAEEAQKAEAAALKKKAEAAEAAAEAAKAAAEKAAADKKAEAEKAAAEKAA 232

RESULT 3  
G85576  
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Nov-2001  
A:Accession: G85576  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; UID:21074935; PMID:11206551  
A:Accession: G85576  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-394 <STO>  
A:Cross-references: GB:AE005174; NID:G12513672; PIDN:AAG55075.1; GSPDB:GN00145; UWGP:209  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: tolA

Query Match 35.0%; Score 181.5; DB 2; Length 394;  
Best Local Similarity 52.2%; Pred. No. 4.7e-05;  
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

QY 1 AKYAKKAE----KAYAKKAAKAEKKAYAKKEAKKAKAEAKKKAKAEAKKYAK-BAAK 55  
DB 120 AEEAAQAEELKQKQAEAAKAAADAKAKAEADDKA--AEAAKAAADAKKAEAEAAK 177  
QY 56 AKKEAY-KAEAKKYAKAKAEKKEVAAAEAKK---AEAAKAYKAEAAKAAAEKAA 106  
DB 178 AAEEAQKAEAAALKKKAEAAEAAAEAAKAAAEKAAAEKAAADKKAEAEKAAAEKAA 232

RESULT 4  
A25550  
histone H1 - sea urchin (Lytechinus pictus)  
C:Species: Lytechinus pictus (painted urchin)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999  
A:Accession: A25550  
R:Knowles, J.A.; Childs, G.J.  
Nucleic Acids Res. 14, 8121-8133, 1986  
A:Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus an

A:Reference number: A25550; UID:87040778; PMID:3022245  
A:Accession: A25550  
A:Molecule type: DNA  
A:Residues: 1-210 <KNO>  
A:Cross-references: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617  
C:Superfamily: histone H1  
C:Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 34.9%; Score 181; DB 2; Length 210;  
Best Local Similarity 50.0%; Pred. No. 3.2e-05;  
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4;

QY 2 KKYAKKAEKAYAKKAAKAEKKAYAKKEAKYAKAAE-----AKKAKAEAKKYAEAAKAK 57  
DB 96 KTEAQKA-RAAAKKAALAAKKKEQKKAATKARKEKLAAKKAAKAAKVKVPPAAKAK 154  
QY 58 KEAYKAEKKYAKAEKAEKKEVAAAEAKKA-----EAAKAYKAEAAKAAAEKAA 106  
DB 155 KPAKKAAPKPAK--KAAKPPAAKPPAKKAAKPPAAKAAKPAKKAAPKAAKPPAAKAA 208

RESULT 5  
E83525  
TolA protein PA0971 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
A:Accession: E83525  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; UID:20437337; PMID:10984043  
A:Accession: E83525  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-347 <STO>  
A:Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04360.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: tolA; PA0971

Query Match 34.5%; Score 179; DB 2; Length 347;  
Best Local Similarity 49.1%; Pred. No. 6.1e-05;  
Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;

QY 6 KKAEEKYAKKAAKAEK-----KAYAKKAEKAYKAAEAKKKAKAEAKKYAEAAKAKE 59  
DB 99 OKLEQQVAAAKAAEQKKADEARKAEQAQAAEAKKAEKAAEAKAAEQKQADIKKR 158  
QY 60 AYKAEAKKYAKAEKKEVAAAEAKKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 109  
DB 159 A-EDEAKK--KAAEDAKKK--AAEDAKKKAAAEAKKAAAEAAKKAAVEA 204

RESULT 6  
AG0592  
tolA protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18  
C:Species: Salmonella enterica subsp. enterica serovar typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
A:Accession: AG0592  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
; S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AG0592  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <PAR>

Best Local Similarity 46.2%; Pred. No. 0.0005;  
Matches 54; Conservative 11; Mismatches 37;  
Indels 15; Gaps 4;



Search completed: March 10, 2003, 12:28:30  
Job time : 32.9641 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 24.8354 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-7

Perfect score: 519

Sequence: 1 AKYAKAEKAYAKAKAAK.....AKAYKAEAKAAKAEAYEA 109

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description
1	189.5	36.5	421	1 TOLA_ECOLI
2	181	34.9	210	1 H1_LYTP1
3	179	34.5	347	1 TOLA_PSEAE
4	175.5	33.8	344	1 MST1_DROHY
5	171.5	33.0	1391	1 MST2_DROHY
6	170.5	32.9	248	1 H1_PARAN
7	169.5	32.7	211	1 H15_STRPU
8	165	31.8	219	1 H10_CHICK
9	163.5	31.5	384	1 TMPB_TREPH
10	163	31.4	224	1 H11L_CHICK
11	160.5	30.9	206	1 H1L_ONCWY
12	160.5	30.9	225	1 H15_HUMAN
13	160	30.8	208	1 H15_WCSM
14	159.5	30.7	218	1 H15_CHICK
15	158	30.4	218	1 H14_HUMAN
16	157	30.3	771	1 CALD_CHICK
17	156.5	30.2	218	1 H101_CHICK
18	154.5	29.8	240	1 H12_VOLCA
19	153.5	29.6	372	1 TOLA_HAETN
20	152.5	29.4	223	1 H103_CHICK
21	151	29.1	1403	1 YDF3_SCHPO
22	150	28.9	211	1 H12_MOUSE
23	150	28.9	217	1 H1G_STRPU
24	148	28.5	139	1 ASH_KLEPN
25	148	28.5	217	1 H1_AANPL
26	147.5	28.4	218	1 H14_MOUSE
27	147	28.3	207	1 H11_CABEL
28	147	28.3	220	1 H13_MOUSE
29	146	28.1	217	1 H1_CHICK
30	146	28.1	299	1 RL22_DROME
31	144	27.7	232	1 H12_GLYBA
32	144	27.7	352	1 ALGP_PSEAE
33	143.5	27.6	218	1 H12_RAT

34	143	27.6	214	1 DBH_MYCTU
35	142.5	27.5	221	1 H1C_CHITE
36	142	27.4	194	1 H1_SALTR
37	141.5	27.3	232	1 H1B_CHITE
38	141.5	27.3	345	1 RS6_AEDAE
39	140	27.0	232	1 H1A_CHITE
40	139	26.8	213	1 H13_RABIT
41	139	26.8	226	1 BASP_HUMAN
42	139	26.8	260	1 H11_VOLCA
43	138.5	26.7	1020	1 NPH_HUMAN
44	138	26.6	235	1 H1E_CHIPA
45	138	26.6	237	1 H1E_CHITE

## ALIGNMENTS

RESULT 1  
TOLA\_ECOLI  
ID TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE TOLA protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JMI05;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levengood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
their products, components of a multistep translocation system in  
Escherichia coli.";  
RL J. Bacteriol. 171:6600-6609(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiuchi T.;  
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levengood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "TolA: a membrane protein involved in colicin uptake contains an  
extended helical region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
RN [5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

```

RX MEDLINE=87040778; PubMed=3022245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RL Lytechinus pictus and Strongylocentrotus purpuratus.";
RL Nucleic Acids Res. 14:8121-8133(1986).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04488; CAA281177.1; -.
CC PIR; A25550; A25550.
CC DR HSP; P02259; 1HST.
CC DR InterPro; IPR001386; Histone H1/H5.
CC DR InterPro; IPR003216; Linkerhist N.
CC Pfam; PF00538; linker histone; 1.
CC DR ProDom; PD000373; Linkerhist N; 1.
CC DR SMART; SM00526; H15; 1.
CC KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
CC SEQUENCE 210 AA; 21746 MW; 08C38F64894007E2 CRC64;
CC -----
Query Match 34.9%; Score 181; DB 1; Length 210;
Best Local Similarity 50.0%; Pred. No. 1.3e-05;
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4;

QY 2 KYAKKAEKAYAKAAAEKKAYAKKAYAKAE-----AKKAAEAKYAKAEAAKAK 57
DB 96 KTAQAKA-RAAAKAKALAAKKKQKQKKAATKARKEKLAAKKAAKAAKKVKPKAAK 154
QY 58 KEAYKAEKAYAKAAAEKKEAYAAAEAKKA-----EAAKAYKAEAAKAAAEKAA 106
DB 155 KPAAKAAKPAAK--KAAKPAAKKPAKAAKPAKAAKPAKAAKPAKAAKPAKAA 208

RESULT 3
TOLA PSEAE
ID TOLA_PSEAE STANDARD; PRT; 347 AA.
AC P50670;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tola protein.
GN TOLA OR PA0971.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=9711325; PubMed=8955385;
RT Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT "Identification and characterization of the tolA genes of
RT Pseudomonas aeruginosa.";
RT J. Bacteriol. 178:7059-7068(1996).
RN [2]
RP REVISIONS TO N-TERMINUS.
RA Duan K., Sokol P.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

```













```
DR EMBL; M17020; AAA48790.1; -.
DR PIR; C28456; C28456.
DR HSP; P08287; IGHC.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone_1.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; DNA-binding; Multigene family.
FT INIT MET 0
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 25 25 METHYLATION (PARTIAL).
FT DOMAIN 37 110 GLOBULAR.
SQ SEQUENCE 218 AA; 21672 MW; CB9724BFF14654A6 CRC64;

Query Match 30.7%; Score 159.5; DB 1; Length 218;
Best Local Similarity 47.7%; Pred. No. 0.0003;
Matches 52; Conservative 11; Mismatches 41; Indels 5; Gaps 3;

QY 1 AKYAKKAEKAYAKAAKAEKKYAKK---EAKYAAEAKKAAKAEAKYAKAEAKAK 57
: || : ||| ||| ||| ||| ||| : ||| : ||| : ||| : |||
Db 109 SKKPGEGLEKAPKKASAAKPKAAAKPAAAKPKKAVAVKSPKKAKKPAASATKKS 168
QY 58 KEAYKAEAK-KYAKAAKAEKKEYAA-AEAKKAEAAKAYKAEAAKAAKE 104
: || : ||| ||| ||| : ||| : ||| : ||| : ||| : |||
Db 169 VKSPKAAKPKKAVAKSPAKAKAVKPKAAKPKAAKPKAAKAAKAAK 217

RESULT 15
H14_HUMAN
ID H14_HUMAN STANDARD; PRT; 218 AA.
AC P10412;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H1.4 (Histone H1b).
GN H1F4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=87057092; PubMed=3782055;
RA Ohe Y., Hayashi H., Iwai K.;
RT "Human spleen histone H1. Isolation and amino acid sequence of a main
RT variant, H1b."
RL J. Biochem. 100:359-368(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9200931; PubMed=1916825;
RA Albig W., Kardalinos E., Drabant B., Zimmer A., Doenecke D.;
RT "Isolation and characterization of two human H1 histone genes within
RT clusters of core histone genes."
RL Genomics 10:940-948(1991).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: THIS VARIANT ACCOUNTS FOR 60% OF HISTONE H1.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M60748; AAA63187.1; -.
DR PIR; A24413; HSHU1B.
DR PIR; C40335; C40335.
DR HSP; P08287; IGHC.
DR Genew; HGNC:4718; H1F4.
DR MIM; 142220; -.
```

```
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone_1.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; DNA-binding; Multigene family.
KW Acetylation; Methylation.
FT INIT MET 0
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 25 25 METHYLATION (PARTIAL).
FT DOMAIN 35 113 GLOBULAR.
SQ SEQUENCE 218 AA; 21734 MW; E210BEC152ABE882 CRC64;

Query Match 30.4%; Score 158; DB 1; Length 218;
Best Local Similarity 48.6%; Pred. No. 0.00038;
Matches 53; Conservative 7; Mismatches 43; Indels 6; Gaps 3;

QY 2 KKYAKKAEKAYAKAAKAEKK--AYAKKAEKAYKAAEAKKAKAEAKYAKAEAKAKE 59
: || : ||| ||| ||| ||| ||| : ||| : ||| : ||| : |||
Db 108 KKAASGEAKPKAKKAGAKAKPKKATGAATPKKSATKTPKAKCPAAAA--G 165
QY 60 AYKAEAKKYAKAAKAEK--KEYAAAAEAKAEAAKAYKAEAAKAAKEAA 106
: || : ||| ||| ||| : ||| : ||| : ||| : ||| : |||
Db 166 AKKAKSPKAKAAKPKKAKPKSPAKAKAVKPKAAKPKTAKPKAAKPKKAA 214
```

Search completed: March 10, 2003, 12:17:11  
Job time : 25.8354 secs



RT	"Evolutionary relationships of conserved cysteine-rich motifs in adhesive molecules of malaria parasites."
RL	Mol. Biol. Evol. 0:0-0(2002).
DR	EMBL; AY042083; AAL10508.1; -.
SQ	SEQUENCE 1866 AA; 212420 MW; DC692D7CFAE7D93F CRC64;
 Query Match 36.7%; Score 190.5; DB 5; Length 1866; Best Local Similarity 49.6%; Pred. No. 4.7e-05; Matches 61; Conservative 17; Mismatches 22; Indels 23; Gaps	
QY	5 AKAEKAYAKKAKEA---KKAVAKKEAKYAAEAACKA-----KAEAKYA 50       :   :   :::     :    :    :    :    :    :    :    :
Db	1262 AKAE--ARKAEAKAEARAKAEAAKKAARAKAEAAKAEARAKAEARAKA 1319 :  :
QY	51 KEAAKAKKEAYKAEAKKYAKAKAEKKEYAAAAEKAEAAK----AYKAAAKAAAEAA 106 :  :
Db	1320 EDARKA-EEARAKAEAKAEARAKAEAKAEAKAEAKAEAKAEAKAEAKA--EAA 1376 :  :
QY	107 YEA 109 :
Db	1377 RKA 1379 :
 RESULT 4 Q8ZQT6 PRELIMINARY; PRT; 407 AA. ID Q8ZQT6; AC Q8ZQT6; DT 01-MAR-2002 (TReMBLrel. 20, Created) DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update) DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update) DB Tol protein, membrane spanning protein. GN TOLA OR STM0747. OS Salmonella typhimurium. OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; OC Salmonella. ON NCBI_TaxID=602; RX [1] RP SEQUENCE FROM N.A. RC STRAIN=UT2 / SGSC1412 / ATCC 700720; RX MEDLINE=21534948; PubMed=11677609; RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.; RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."; RL Nature 413:852-856(2001). DR EMBL; AF008730; AAL19691.1; -. DR InterPro; IPR000104; Antifreeze_1. DR PRINTS; PR00308; ANTIFREEZEI. KW Complete proteome. SQ SEQUENCE 407 AA; 41865 MW; 2534352116602D75 CRC64;	
 Query Match 35.5%; Score 184.5; DB 16; Length 407; Best Local Similarity 55.5%; Pred. No. 3e-05; Matches 66; Conservative 11; Mismatches 25; Indels 17; Gaps	
QY	1 AKKYAKKAEKAVAKKAAKKEKAYAKKEAKYK-AEAKKKKAEAKKYAKGA-AKAKK 58     :    :    :    :    :    :    :    :    :    :    :    :    :
Db	138 AAAADAKKKAEAEAKAAADA-KKAEAEAVKAAADAKKKAEAEAKAAADAKKAAEA 195     :    :    :    :    :    :    :    :    :    :    :    :    :
QY	59 EAYK--AEAKKYAKKAEKKEYAAAEEK---AEAKYKAEAAK---AAAKEAAVEA 109     :    :    :    :    :    :    :    :    :    :    :    :    :
Db	196 EAATAAAEAKKKAEAAK----AAAEAKKADAEAAKA-AAEANKKDAARAATAAAEA 249     :    :    :    :    :    :    :    :    :    :    :    :    :
 RESULT 5 IX8X965 PRELIMINARY; PRT; 394 AA. ID Q8X965; AC Q8X965; DT 01-MAR-2002 (TReMBLrel. 20, Created) DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)	



DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Membrane spanning protein, required for outer membrane integrity  
 GN TOLA OR Z0907 OR EC50774.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL; AE005252; AAG55075.1; -.  
 DR EMBL; AP002553; BAB34197.1; -.  
 DR InterPro; IPR000104; Antifreeze\_1.  
 DR PRINTS; PR00308; ANTIFREEZEI.  
 KW Complete proteome.  
 SQ SEQUENCE 394 AA; 40517 MW; 5858D8E8230BDE28 CRC64;

Query Match 35.0%; Score 181.5; DB 16; Length 394;  
 Best Local Similarity 52.2%; Pred. No. 4.5e-05;  
 Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

Oy 1 AKKYAKKAE-----KAYAKKAAKKKAYAKKAEKAYKAAEKKKAKAEKKYAK-EAAK 55  
 Db 120 ABEAAKQAEKQKQAEAAKAAADAKAKAEADDKA--AEAAKAAADAKKKAEAAK 177  
 Oy 56 AKKEAY-TAEAKKYAKAAEKKKAYAAEAKK--AEAAKAYKAAEAAKAAEAA 106  
 Db 178 AAEEAQKAAEAAALKKKAAEAAEAAAEAAEARKKAAEAAEAAADKKAAEAAEAA 232

RESULT 6  
 ID Q828C1 PRELIMINARY; PRT; 376 AA.  
 AC Q828C1;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Tola protein.  
 GN ST0793.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 enterica serovar Typhi CT18."  
 RL Nature 413:848-852(2001).  
 DR EMBL; AL627288; CAD05209.1; -.  
 DR InterPro; IPR000104; Antifreeze\_1.  
 DR PRINTS; PR00308; ANTIFREEZEI.  
 DR PRINTS; PR01574; TUBBYPROTEIN.  
 KW Complete proteome.  
 SQ SEQUENCE 376 AA; 38804 MW; EC21F2C4767A8A42 CRC64;

Query Match 34.3%; Score 178; DB 16; Length 376;  
 Best Local Similarity 52.3%; Pred. No. 7.4e-05;  
 Matches 58; Conservative 10; Mismatches 33; Indels 10; Gaps 4;

Oy 1 AKKYAKKAEKAYAKKAEKKKAYAKKAEKAYKAAEKKKAKAEKKYAKAEAAKKEA 60  
 Db 137 AKAADADAKKKAEAAEAAKAAADAKKKAEAEA-AKAAADAKKKAEAEA---AKAAEAKK-- 190  
 Oy 61 YKAEAKKYAKAAEKKKAEYAAAEAAEAAKAEAAKAYKAAEAAKAAA---KEAAEY 108  
 Db 191 -KAEAAEAAKAAADAKKKADAEAAKAEAAEKKKADAAAKAAADAKKKAAAE 240

RESULT 7  
 ID Q8ZGZ2 PRELIMINARY; PRT; 388 AA.  
 AC Q8ZGZ2;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Tola colicin import membrane protein.  
 GN TOLA OR YP01123.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Kariyeh A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
 RL Nature 413:523-527(2001).  
 DR EMBL; AJ414146; CAC89966.1; -.  
 DR InterPro; IPR000104; Antifreeze\_1.  
 DR PRINTS; PR00308; ANTIFREEZEI.  
 KW Complete proteome.  
 SQ SEQUENCE 388 AA; 40424 MW; 81447B04B30A7E7C CRC64;

Query Match 34.3%; Score 178; DB 16; Length 388;  
 Best Local Similarity 52.1%; Pred. No. 7.6e-05;  
 Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;

Oy 2 KKYAKKAEKAYAKKAEKKK-----AYAKKEA-KAYKA-AEAKKKAKAEKKYAKKEA 53  
 Db 140 KQAEQQKIAAANAVAKAEKEQKQAEATAAQAKAEADKIVKQAEQKAEAEKAEAAVA 199  
 Oy 54 AKAKKEAYKAEAKKYAKAAK-----AEKKEAYAAAEAKKAEAAK--AYKAEAAKAAKKEA 105  
 Db 200 AAANKQA-DADAKKAVEAEKAAADAEEKKAAADAE-KKAAAKKVAANAAEAKKAAAEA 257  
 Oy 106 A 106  
 Db 258 A 258

```
RESULT 8
Q8XVW7
ID Q8XVW7 PRELIMINARY; PRT; 200 AA.
AC Q8XVW7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Probable histone H1 protein.
GN RSC2793 OR RS00453.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646071; CAD16500.1; -.
DR InterPro; IPR00104; Antifreeze_1.
DR InterPro; IPR003993; treacle.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01503; TREACLE.
KW Complete proteome.
SQ SEQUENCE 200 AA; 19279 MW; D3831B590510272D CRC64;

Query Match 34.0%; Score 176.5; DB 16; Length 200;
Best Local Similarity 54.5%; Pred. No. 5.2e-05;
Matches 67; Conservative 5; Mismatches 32; Indels 19; Gaps 8;

QY 1 AKYAKAEKAYAKAAKAEK---KKAYAKAEKAYAKAAK---KKA----KAEAKY- 49
Db 42 AKKVA--AKKVAAKKAPAAKAAKAAVKKVAAKAAKAAKAAKAAKAAKAAKAAKVA 99
QY 50 AKEAAKAEKAYAKAAKAAK---AEKKEAYAAAEKAEKAAKAYKAE-AAKAAAEKAA 106
Db 100 AKGAAPAKAAVKKVAAKAAKAAKAAK---APAKKAPAKAAKAAKAAKAAKAAKPA 156
QY 107 YEA 109
Db 157 KKA 159

RESULT 9
Q9CMW70
ID Q9CMW70 PRELIMINARY; PRT; 389 AA.
AC Q9CMW70;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ToLA.
GN TOLA OR PM0968.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMW70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; A006136; AAK03052.1; -.
```

```
DR HSP; P19934; 1TOL.
DR InterPro; IPR00533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Complete proteome.
SQ SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9E94B CRC64;

Query Match 33.4%; Score 173.5; DB 16; Length 389;
Best Local Similarity 48.7%; Pred. No. 0.00015;
Matches 55; Conservative 18; Mismatches 31; Indels 9; Gaps 4;

QY 6 KKAEEKAYAK-AAAEEK-KAYAKKEAKAYAKAAKKA--EAKYAKEA-----AKA 56
Db 145 KQAEERAKQLAEAAKAKAEAEKRLAALAKQAEERAKAEAEKAKAEAEKAKAEABAKA 204
QY 57 KKEAYKAEAKYAKAAKAEKKEAYAAAEAKKAEAAKAYKAEAAKAAKAAKAAVEA 109
Db 205 KVEKAKAEAEAKVKAEKAKAEAEKAKAEAKAEKAKAEKAKAEKAKADAAQAORKA 257

RESULT 10
O61164
ID O61164 PRELIMINARY; PRT; 1701 AA.
AC O61164;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE Erythrocyte binding protein.
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria
RT parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235(1998).
DR EMBL; AF031886; AAC05366.1; -.
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87CE8A CRC64;

Query Match 32.9%; Score 170.5; DB 5; Length 1701;
Best Local Similarity 53.5%; Pred. No. 0.00091;
Matches 54; Conservative 11; Mismatches 31; Indels 5; Gaps 5;

QY 5 AKKAEKAYAKAAKAEKKAYAKKEAKAYAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA- 63
Db 1205 AKKAEK-RRKAEAVK-KAEAEKAAKAAKAAKAEERKK-KAEAAKALERKKKSEAKKAL 1261
QY 64 EAKYAKAAKAEKKEAYAAAEAKKAEAAKAYKAEAAKAAKAAKAAKAAKAAKAAKAAKE 104
Db 1262 ERKKAEAAKAAKAEKKKAEAAKAAKAAKAESEKK-KAEAAKAAKAAKEE 1301

RESULT 11
Q39576
ID Q39576 PRELIMINARY; PRT; 232 AA.
AC Q39576;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Histone H1.
GN CHI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Muller K., Lindauer A., Park P.B., Cornelius T., Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas
```

RT histone genes reveal features linking plant and animal genes."

```
RL Curr. Genet. 28:333-345(1995).
DR EMBL; U16726; AAA98452.1; -.
DR HSSP; P02259; 1HST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; Linkerhist N.
DR Pfam; PF00538; linker_histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
SQ SEQUENCE 232 AA; 24693 MW; 2D006AE44A8FA037 CRC64;

Query Match 32.8%; Score 170; DB 10; Length 232;
Best Local Similarity 46.3%; Pred. No. 0.00016;
Matches 57; Conservative 10; Mismatches 38; Indels 18; Gaps 5;

QY 2 KKYAKA-----EKAYAKAKAAKEKKAYAKEA-----KAYKAAEAKKKAKAAEKYAKAE- 53
Db 100 KKAACKAAPTAKAAAPKKEGAVKTKAPKKEGKPKSAKKAKEKPKGEGKKAAKPAK 159

QY 54 -----AKAKEAYKAEAKKYAKAKAAKEKKEAYAAE-----AKKAAEAK---AYKAAEAKAAAK 103
Db 160 ABEKKPAAPAKPTPTTKAAAPKAEKPKKAAAKPAEKPKPAKPAEKPAEKPAEKPAKPAK 219

QY 104 EAA 106
Db 220 KSA 222

RESULT 12
CO01395 PRELIMINARY; PRT; 275 AA.
ID CO01395;
AC CO01395;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Axoneme-associated protein MST101(3).
GN MST101(3) OR DHMST101.
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A.
RA Neesen J., Heinlein U.A.O., Buenemann H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC -!- SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR EMBL; U85627; AAB51369.1; -.
DR FlyBase; FBgn0020732; Dhyd\mst101(3).
KW Sperm; Repeat; Multigene family.
FT DOMAIN 64 255
FT 13 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-X-
FT [AE]-X.
SQ SEQUENCE 275 AA; 30436 MW; 76BA7B2A2DF732C CRC64;

Query Match 32.7%; Score 169.5; DB 5; Length 275;
Best Local Similarity 48.2%; Pred. No. 0.0002;
Matches 54; Conservative 16; Mismatches 29; Indels 13; Gaps 6;

QY 1 AKKYAKKAEKAYAKK-AKAAEKKAYAKAEKAYKAAEAKKKAK-AEAKKYAKAEAKAK 57
Db 51 AEDVKKCEAAKKAACKAAEKKAEKAEKK-----KCAEAAKKEKAEAKKKCAEAAKKE 106

QY 58 KEAYKAEAKKYAKAKAEKKEKAYAAEAKK-AEAAKAYKAAEAAKAAEAYE 108
Db 107 QEA--AQKKCAELAKKEK-----AEKKKCAEAAKKEAEAKKKCEAAFK 153

RESULT 13
```

```
Q95S18 PRELIMINARY; PRT; 619 AA.
ID Q95S18;
AC Q95S18;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE HL01392p.
GN FAU OR CG6544.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzales M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060997; AAL28545.1; -.
DR FlyBase; FBgn0020439; fau.
SQ SEQUENCE 619 AA; 68641 MW; 7DDCB26AD1AB9CEE CRC64;

Query Match 32.1%; Score 166.5; DB 5; Length 619;
Best Local Similarity 41.0%; Pred. No. 0.00067;
Matches 55; Conservative 16; Mismatches 30; Indels 33; Gaps 4;

QY 9 EKAYAKKAAKAEKKAYAKAEKAAEAKKKAKAAEAKKKYAKAEAK-AKKEAYKAE--A 65
Db 331 EKKRAQKADEAKREERALKERDLRTAEAEKQAAKAKAAEAAKTAABEALLAEAAA 390

QY 66 KKYAKAAK-----EKKEVAAAEAKKAAEAKKAYKA-- 95
Db 391 OKAAEAKALKAAEDAAOKAAAEARLABEAAAKQVAEEAAKAAEAEARLAEAAKAAE 450

QY 96 EAAKAAAEKAAAYEA 109
Db 451 EAAQKAAEAAKLA 464

RESULT 14
Q39598 PRELIMINARY; PRT; 265 AA.
ID Q39598;
AC Q39598;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Cgcr-4 product (Fragment).
GN CGCR-4.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119224; PubMed=1731966;
RA Wakarchuk W.W., Muller F.W., Beck C.P.;
RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
RT arrangements of directly repeated sequence motifs.";
RL Plant Mol. Biol. 18:143-146(1992).
DR EMBL; X17208; CAA35080.1; -.
FT NON TER 1
SQ SEQUENCE 265 AA; 26216 MW; B35318B777CF782 CRC64;

Query Match 31.6%; Score 164; DB 10; Length 265;
Best Local Similarity 39.4%; Pred. No. 0.00045;
Matches 50; Conservative 21; Mismatches 38; Indels 18; Gaps 2;

QY 1 AKKYAKKAEKA-----YAKKAAEKKKAYAKKAEKAYKAAEAKKKAKAE 46
Db 1 AKKYAKKAEKA-----YAKKAAEKKKAYAKKAEKAYKAAEAKKKAKAE 46
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 56.7996 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-7  
Perfect score: 519  
Sequence: 1 AKYAKAEKAYAKAKAAK.....AKYAKAEKAAKAAKAAEAYEA 109

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 13250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	519	100.0	109	AA1982577	Copolymer molecule
2	312.5	60.2	86	AA1982576	Copolymer molecule
3	289	55.7	77	AA1982575	Copolymer molecule
4	228.5	44.0	66	AA1982574	Copolymer molecule
5	209	40.3	154	AA1982573	Recombinant copoly
6	187	36.0	106	AA1982572	Recombinant copoly
7	180.5	34.8	56	AA1982571	Copolymer molecule
8	179.5	34.6	100	AA1982570	Peptide #10 used i
9	179.5	34.6	100	AA1982569	Amino acid polymer
10	179.5	34.6	100	AA1982568	Poly-Lys-Ala used

11	179.5	34.6	100	22	AA1982567	Nucleic acid trans
12	178	34.3	223	20	AA1982566	Amino acid sequenc
13	162.5	31.3	643	16	AA1982565	Trypanosoma cruzi
14	160.5	30.9	158	20	AA1982564	Histone H1 isoform
15	160.5	30.9	158	21	AA1982563	Human histone H1.5
16	160.5	30.9	226	20	AA1982562	Human histone H1.1
17	160.5	30.9	226	21	AA1982561	Human histone H1.1
18	159.5	30.7	234	23	AA1982560	Human histone H1.1
19	159.5	30.7	234	23	AA1982559	Human linker histone
20	158.5	30.5	472	16	AA1982558	Trypanosoma cruzi
21	158	30.4	218	18	AA1982557	Human histone H1.1
22	158	30.4	218	20	AA1982556	Human histone H1.1
23	158	30.4	218	20	AA1982555	Human histone H1.1
24	157.5	30.3	564	16	AA1982554	Trypanosoma cruzi
25	154.5	29.8	130	23	AA1982553	Peptide fragment o
26	154.5	29.8	130	23	AA1982552	Human linker histone
27	154	29.7	219	21	AA1982551	Quail H1 histone p
28	152.5	29.4	222	18	AA1982550	Human histone H1.1
29	152.5	29.4	222	20	AA1982549	Human histone H1.1
30	152.5	29.4	222	21	AA1982548	Human histone H1.1
31	151	29.1	607	22	AA1982547	Drosophila melanog
32	150	28.9	1507	21	AA1982546	Plasmodium yoelii
33	148.5	28.6	140	13	AA1982545	Synthetic helical
34	146	28.1	299	22	AA1982544	Drosophila melanog
35	144	27.7	111	18	AA1982543	Adhesive polyptei
36	144	27.7	434	18	AA1982542	Antibiotic potent
37	144	27.7	434	21	AA1982541	(Lys)434 protein s
38	143	27.6	214	20	AA1982540	M. tuberculosis hi
39	143	27.6	214	21	AA1982539	M. tuberculosis hi
40	141.5	27.3	239	23	AA1982538	Listeria monocytog
41	141	27.2	100	21	AA1982537	Peptide #8 used in
42	141	27.2	100	21	AA1982536	Amino acid polymer
43	141	27.2	100	22	AA1982535	Poly-L-Lysine used
44	141	27.2	100	22	AA1982534	Nucleic acid trans
45	138.5	26.7	617	22	AA1982533	Peptide #2951 enco

ALIGNMENTS

RESULT 1  
AA1982577  
ID AA1982577 standard; peptide; 109 AA.  
XX AA1982577;  
XX AC  
XX AA1982577;  
DT 28-JUL-2000 (first entry)  
XX  
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:7.  
XX  
KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX Unidentified.  
XX OS  
XX WO200018794-A1.  
XX  
PD 06-APR-2000.  
XX  
PF 24-SEP-1999; 99WO-US22402.  
XX  
PR 25-SEP-1998; 98US-0101693.  
XX  
PA (YEDA ) YEDA RES & DEV CO LTD.  
PA (TEVA-) TEVA PHARM USA INC.





Db 26 -KAEAKKYAKAAKAEKKYEAAYEA-----YKAEAAKAAKAEAYEA 66

RESULT 5  
AAR06445  
ID AAR06445 standard; protein; 154 AA.  
XX AAR06445;

XX 03-JAN-1991 (first entry)

XX Recombinant copolymer 1-77, myelin basic protein analogue.

XX Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
XX immunological activity; autoimmune encephalomyelitis;  
XX multiple sclerosis;

XX Synthetic.

XX EP383620-A.

XX 22-AUG-1990.

XX 16-FEB-1990; 90EP-0301700.

XX 07-FEB-1990; 90US-0473845.

XX 17-FEB-1989; 89US-0312541.

XX (REPL-) REPLIGEN CORP.

XX Cook KS;

XX WPI; 1990-255848/34.

XX N-PSDB; AAQ05664.

XX Producing genes encoding random polymers of aminoacid(s) - for

XX producing recombinant polypeptide(s) with biological and/or

XX immunological activity

XX Disclosure; Fig 11; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes  
XX coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaTAN  
XX (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
XX express Protein A. The resulting plasmids encode fusion proteins  
XX consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
XX A methionine residue occurs between the Protein A and rCOP-1  
XX sequences, originating from the 5' linker sequence, in order that  
XX the COP-1 polypeptide may be cleaved from the fusion protein.  
XX rCOP-1-77 contains oligonucleotide duplexes encoding the following  
XX segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue  
XX is left behind following CNBr cleavage of the fusion protein.

XX The product prevents or arrests experimental autoimmune  
XX encephalomyelitis. They are used to prevent, arrest or control a  
XX demyelinating disorder, e.g. multiple sclerosis. They may also  
XX be used as additives to hair care products to confer beneficial  
XX effects on damaged hair or as supplements for diets deficient in  
XX certain amino acids.  
XX See also AAQ05665.

XX Sequence 154 AA;

Query Match 40.3%; Score 209; DB 11; Length 154;  
Best Local Similarity 53.7%; Pred. No. 1.6e-10;  
Matches 72; Conservative 8; Mismatches 20; Indels 34; Gaps 9;

QY 2 KKYAKKAEKA-----YAKKAKAAKKEKKAYA----KKEAKAYKAAEA-----KKKAK-A 44  
DB 24 KKKAKAEKAKKAKYKKYKKEAEAAKAAKAAKAAAYKKAEAE--AAEAEKAKYKKKAKAE 81

QY 45 EAKKYAEKAEKA--KKEAYKAEKKYKAKAKAEK-----KEYAAAEKAEAA--K 91

DB 82 EYKCKAKAAAEAYKKAEAEYKKYKKKAKKAKYKKKAEKAKAKAAAEKAEKAEYK 141

QY 92 AYKAEAAKAAKAEK 105  
DB 142 KYKKEAEK--AKEA 153

RESULT 6  
AAR06446  
ID AAR06446 standard; protein; 106 AA.

XX AAR06446;

XX 03-JAN-1991 (first entry)

XX Recombinant copolymer 1-19, myelin basic protein analogue.

XX Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;  
XX immunological activity; autoimmune encephalomyelitis;  
XX multiple sclerosis;

XX Synthetic.

XX EP383620-A.

XX 22-AUG-1990.

XX 16-FEB-1990; 90EP-0301700.

XX 07-FEB-1990; 90US-0473845.

XX 17-FEB-1989; 89US-0312541.

XX (REPL-) REPLIGEN CORP.

XX Cook KS;

XX WPI; 1990-255848/34.

XX N-PSDB; AAQ06446.

XX Producing genes encoding random polymers of aminoacid(s) - for

XX producing recombinant polypeptide(s) with biological and/or

XX immunological activity

XX Disclosure; Fig 12; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes  
XX coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaTAN  
XX (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
XX express Protein A. The resulting plasmids encode fusion proteins  
XX consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
XX A methionine residue occurs between the Protein A and rCOP-1  
XX sequences, originating from the 5' linker sequence, in order that  
XX the COP-1 polypeptide may be cleaved from the fusion protein.  
XX rCOP-1-19 contains oligonucleotide duplexes encoding the following  
XX segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA KEA, and AAA. The  
XX N-terminal alanine residue is left behind following CNBr cleavage of the  
XX fusion protein.

XX The product prevents or arrests experimental autoimmune  
XX encephalomyelitis. They are used to prevent, arrest or control a  
XX demyelinating disorder, e.g. multiple sclerosis. They may also  
XX be used as additives to hair care products to confer beneficial  
XX effects on damaged hair or as supplements for diets deficient in  
XX certain amino acids.  
XX See also AAQ05664.

XX Sequence 106 AA;

Query Match 36.0%; Score 187; DB 11; Length 106;  
Best Local Similarity 55.0%; Pred. No. 7.7e-09;  
Matches 60; Conservative 11; Mismatches 20; Indels 18; Gaps 6;

QY 5 AKKAEKAYAKKAAKKEKKAYAKKEAKAYKAAEAKKKKAEAKKAYKAEAKKAEYKAE 64

DB 1 AKAEKAKAAK-KAYEAEKAKAKAYEAK--KAEKAEKAEKAAAEKAEKAEKAEK-- 51





[illegible]





```
RESULT 15
AAV57366
ID AAV57366 standard; peptide; 158 AA.
XX
AC AAV57366;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human histone H1.5 pANCA-reactive fragment (residues 69-226).
XX
KW Ulcerative colitis; inflammatory bowel disease; porin antigen;
KW pANCA; perinuclear anti-neutrophil cytoplasmic antibody; human;
KW histone H1; isoform; NANUC-2.
XX
OS Homo sapiens.
XX
PN US6033864-A.
XX
PD 07-MAR-2000.
XX
PF 12-MAR-1998; 98US-0041889.
XX
PR 12-APR-1996; 96US-0057846.
XX
PR 11-APR-1997; 97US-0837058.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Cohavy O, Braun J;
XX
DR WPI; 2000-255695/22.
XX
PT Diagnosing ulcerative colitis or susceptibility, by detecting complex
PT formation between microbial porin antigen and perinuclear
PT anti-neutrophil cytoplasmic autoantibodies -
XX
PS Example 4; Columns 69-70; 49pp; English.
XX
CC The invention provides a method for diagnosing ulcerative colitis in a
CC subject suspected of having inflammatory bowel disease. The method
CC comprises reacting a patient sample with a porin antigen that is
CC immunologically reactive with pANCA (perinuclear anti-neutrophil
CC cytoplasmic antibodies) and detecting formation of a Ag-pANCA complex
CC as indicative of ulcerative colitis. The method is used to diagnose
CC ulcerative colitis or susceptibility to it. Sequences AAV57362-367
CC represent pANCA-reactive fragments, derived from human histone H1.5.
XX
SQ Sequence 158 AA;

Query Match 30.9%; Score 160.5; DB 21; Length 158;
Best Local Similarity 46.8%; Pred. No. 1.9e-06;
Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

Qy 1 AKVYAKAEKAYAKAKAEKKEKAYAKKEAKYKAA-----EAKKKAKAEAKKYAKEAAK 55
Db 51 AKPKAKKAGAAKAKKPKAGATPKK--AKKAGAKKAVKKTPTKKKPPAAAGVKVAKSPKK 108
Qy 56 AKKEAYKAEKAYAKAEKAEKKEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 104
Db 109 AKAAAKPKKATKSPAKPKAVKPKAKPKAKPKAKPKAKPKAKPKAKPKAKPKAKPKAKPK 157

Search completed: March 10, 2003, 12:21:31
Job time : 57.7996 secs
```

**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	519	100.0	109	10	US-09-816-989A-7		Sequence 7, Appli
2	312.5	60.2	86	10	US-09-816-989A-6		Sequence 6, Appli
3	289	55.7	77	10	US-09-816-989A-5		Sequence 5, Appli
4	228.5	44.0	66	10	US-09-816-989A-4		Sequence 4, Appli
5	180.5	34.8	56	10	US-09-816-989A-3		Sequence 3, Appli
6	178	34.3	223	9	US-10-051-643-201		Sequence 201, App
7	162	31.2	356	9	US-09-820-843A-27		Sequence 27, Appl
8	154.5	29.8	309	9	US-03-820-843A-24		Sequence 24, Appl
9	153.5	29.6	369	9	US-03-820-843A-95		Sequence 95, Appl
10	153.5	29.6	372	9	US-09-820-843A-8		Sequence 8, Appli
11	144	27.7	352	9	US-09-820-843A-23		Sequence 23, Appl
12	144	27.7	434	10	US-09-124-280A-3		Sequence 3, Appli
13	143.5	27.6	452	9	US-10-184-832-5		Sequence 5, Appli
14	138.5	26.7	617	10	US-09-864-761-36182		Sequence 36182, A
15	138	26.6	45	10	US-09-816-989A-2		Sequence 2, Appli
16	126	24.3	600	9	US-09-738-626-5197		Sequence 5197, Ap
17	122.5	23.6	218	9	US-09-999-724-48		Sequence 48, Appl
18	122.5	23.6	829	10	US-09-815-243-11216		Sequence 11216, A
19	122.5	23.6	892	10	US-09-815-243-13765		Sequence 13765, A

## ALIGNMENTS

; Sequence 6, Application US/09816989A

```

; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; TITLE OF INVENTION: Lys, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match          60.2%; Score 312.5; DB 10; Length 86;
Best Local Similarity 72.1%; Pred. No. 1.5e-18;
Matches 80; Conservative 1; Mismatches 3; Indels 27; Gaps 5;

QY 1 AKYAKAEKAYAKGAAKAEKKAYAKKEKAYAKKAAEAKKKAKAEAKKYAKAEAAKKEA 60
DB 1 AKYAKK-EKAYAKK-----EKAAKKAEEKAYAKAEAKKKA----- 36

QY 61 YKAEAKKYAKAAKAEKKEYAAAAEAK-KAEAA-KAYKAEAAKAAKAEAAEAA 109
DB 37 -KAEAKKYAKAAKAEKKEYAAAAEAKYKAEAAKKAYKAEAAKAAKAEAAEAA 86

RESULT 3
US-09-816-989A-5
; Sequence 5, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lys, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match          55.7%; Score 289; DB 10; Length 77;
Best Local Similarity 67.0%; Pred. No. 9.1e-17;
Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps 4;

QY 1 AKYAKAEKAYAKGAAKAEKKAYAKKEKAYAKKAAEAKKKAKAEAKKYAKAEAAKKEA 60
DB 1 AKYAKK-EKAYAKK-----EKAAKKAEEKAYAKAEAKKKA----- 36

QY 62 YKAEAKKYAKAAKAEKKEYAAAAEAK-KAEAA-KAYKAEAAKAAKAEAAEAA 109
DB 37 -KAEAKKYAKAAKAEKKEYAAAAEAKYKAEAAKKAYKAEAAKAAKAEAAEAA 77

```



Db 1 AKYAKK-EKAYAKK-----EKAAKKAEKAYKAAEAKKKAEAK-----39  
Qy 61 YKAEAKKYAKAAKAEKKEKYAAAEAKKAEAAKAYKABAAKAAKAAEAYEA 109  
Db 40 -----YKAEAAKAAKAEAYEA 56

## RESULT 6

US-10-051-643-201  
; Sequence 201, Application US/10051643  
; Publication No. US20020197265A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Tan, Paul L. J.  
; TITLE OF INVENTION: Methods and Compounds for the Treatment  
; TITLE OF INVENTION: Of Immunologically-Mediated Diseases of the Respiratory  
; TITLE OF INVENTION: System using Mycobacterium Vaccae  
; FILE REFERENCE: 11000.1008c2  
; CURRENT APPLICATION NUMBER: US/10/051,643  
; CURRENT FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US09/156,181  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: US 08/996,624  
; PRIOR FILING DATE: 1997-12-23  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 201  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Mycobacterium vaccae  
US-10-051-643-201

Query Match 34.3%; Score 178; DB 9; Length 223;  
Best Local Similarity 54.5%; Pred. No. 1.3e-07;  
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;  
Qy 1 AKYAKK-AEKYAKKAKAAKE-----KKAYAKKEKAYKAAEAKKKAKAEAKKYKAEAAK 55  
Db 112 ARKAAPKAPAKAAKAAKAPAKAPAKKA-ATKAAPAKKATAAKKAAKAPAKKATAAKKAAAP 170  
Qy 56 AKKEAYKAEAKKYKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAK 103  
Db 171 AKKAPAKKAATKAAKAPKAPKAAKATAKAPAKKAPAAK--KAPAKKAPAK 218

## RESULT 7

US-09-820-843A-27  
; Sequence 27, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 27  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Vibrio cholerae  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: tola protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|9656364  
US-09-820-843A-27

Query Match 31.2%; Score 162; DB 9; Length 356;  
Best Local Similarity 42.6%; Pred. No. 3.6e-06;  
Matches 55; Conservative 13; Mismatches 37; Indels 24; Gaps 5;

Qy 3 KYAKKAEKAYAKK--AKAAKEKKAYAKKEAKYKAAEAKKK-----AKAEAKKYAKAEAAK 56  
Db 112 KAAREAEKLREQEQERLAAEQKAREEKERAAKAEAKRVKVEEAAKKAQEERVAKEAAAA 171  
Qy 57 KKEAYKAEAKKYAK-----AKAAEK-----KEVAAAEAKKAEAAKAYKAEAAK 100  
Db 172 KAEQORIEREKAKLAEBEKAKREKEVAKAEQERLAKERAKKEA--ADKAKKEKERAACA 229  
Qy 101 AAKEAAVEA 109  
Db 230 EAERKAQEA 238

## RESULT 8

US-09-820-843A-24  
; Sequence 24, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 24  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: polyhydroxyalkanoate synthesis protein Phaf  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|9951352  
US-09-820-843A-24

Query Match 29.8%; Score 154.5; DB 9; Length 309;  
Best Local Similarity 46.9%; Pred. No. 1.2e-05;  
Matches 53; Conservative 3; Mismatches 50; Indels 7; Gaps 2;  
Qy 1 AKYAKKAEKAYAKKAAKAEKKAYAKKEAK-----AYKAAEAKKKAKAEAKKYKAEAA 54  
Db 150 AKPAKPAKTAATAAKPAKPAKAAKAAKPAKPAKKTAAKTAATAAKPAKPAKPTAKAAA 209  
Qy 55 K-AKKEAYKAEAKKYKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAKAA 106  
Db 210 KPAKPAKAAKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 262

## RESULT 9

US-09-820-843A-95  
; Sequence 95, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 95  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: T. pallidum  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: conserved hypothetical protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|3322751  
US-09-820-843A-95

```

Query Match      27.7%; Score 144; DB 9; Length 352;
Best Local Similarity 44.0%; Pred. NO. 9.2e-05;
Matches 55; Conservative 1; Mismatches 53; Indels 16; Gaps 3;

QY 1 AKKYAKAEKAYAKKAAKEKKAYAKKEAK-----AYKAAEAKKKAKAEAKKYAKEAAK 55
   || || || || || || || || || || || || || || || || || || || ||
Db 164 AKPAAPAAKPAAPAAKTAAPAAKPAAPKPTAKPAAPKPAAPKPAAPKTAAPAAKPAAPKPAAPKPAAPK 223
   || || || || || || || || || || || || || || || || || || || ||
QY 56 -----AKKEAYKAEAKKYAK-----AAKAEKKEYAAAEAKKAEAAKAYKAAEAAKAAAKE 104
   || || || || || || || || || || || || || || || || || || || ||
Db 224 PAAKTAAPAAKPAAPKPAAPKPAAPKPTAKPAAPKTAAPAAKPAAPKPAAPKPAAPKPAAPKPAAPK 283
   || || || || || || || || || || || || || || || || || || || ||
QY 105 AAYEA 109
   || || || || || || || || || || || || || || || || || || ||
Db 284 AAKPA 288

RESULT 12
US-09-124-280A-3
; Sequence 3, Application US/09124280A
; Patent No. US20020034520A1
; GENERAL INFORMATION:
; APPLICANT: Porrio, Massimo
; TITLE OF INVENTION: VACCINES FOR PREVENTION OF GRAM-
; TITLE OF INVENTION: NEGATIVE BACTERIAL INFECTIONS AND ENDOTOXIN RELATED D
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hedman, Gibson & Costigan
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIA TYPE: Diskette, 3.50 inch, 1.44 Mb storage

```

```

COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-3

Query Match      27.7%; Score 144; DB 10; Length 434;
Best Local Similarity 36.9%; P-red. No. 0.00011;
Matches 38; Conservative 14; Mismatches 51; Indels 0; Gaps 0;

QY    2 KKVAKAEKAYAKKAKEKKAYAKEKAYAAEAKKAKAEAKKYAKAEAKKAEAY 61
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db    278 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 337

QY    62 KAFKKYAKAEKAEKYEAAAEEAKKAEAAKAYKAEAAKAAAE 104
      ||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db    338 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 380

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36182
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005529.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
; OTHER INFORMATION: SWISSPROT HIT: P12036, EVALU= 2.00e-33
; OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALU= 3.00e-22
; US-09-864-761-36182

Query Match 26.7%; Score 138.5; DB 10; Length 617;
Best Local Similarity 35.3%; Pred. No. 0.00044;
Matches 49; Conservative 17; Mismatches 36; Indels 37; Gaps

QY 5 AKKAEKAYAK---KAKAAKEKKAYAKKEAYKAAEAKKKAKAEAKKYAKEAAKAK---- 57
||| ||| : : : : : ||| : : : ||| : : : ||| : : : |||
Db 161 AKSPEKEEAKPAEVKSPKAKSPAKEEAKSPAEAKSPKVEAKSPKVEAKSPAEAKSPVK 220
||| ||| : : : : : ||| : : : ||| : : : ||| : : : |||
QY 58 -----KEAYKAAKKYAKAAKAEKKEYYA-----AEAKKAEAKA- 92
||| ||| : : : ||| ||| ||| ||| : : : ||| ||| |||
Db 221 EEAKSPAENVKSPKAKSPPTKEAKSPKAKSPKEAKSPKAKSPKVEAKSPKAKSPK 280
||| ||| : : : ||| ||| ||| ||| : : : ||| ||| |||
QY 93 YKAEA-----AKAAAKEAA 106
||| ||| : : : ||| ||| ||| ||| : : : ||| ||| |||
Db 281 VKAEKSPKAKSPVKEEA 299
||| ||| : : : ||| ||| ||| ||| : : : ||| ||| |||

RESULT 15
US-09-816-989A-2
; Sequence 2, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECUL
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Pepti
; US-09-816-989A-2

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 19.3165 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-7  
Perfect score: 519  
Sequence: 1 AKKYAKAEKAYAKAKAAK.....AKAYKAEKAAKAAKAAVEA 109

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179.5	34.6	100	2	US-08-460-890A-64
2	179.5	34.6	100	3	US-08-167-641C-64
3	179.5	34.6	100	4	US-08-460-971A-64
4	179.5	34.6	100	4	US-08-462-040-64
5	178	34.3	223	4	US-09-095-855-201
6	178	34.3	223	4	US-09-205-426-201
7	162.5	31.3	643	2	US-08-216-894-8
8	162.5	31.3	643	4	US-09-115-746-8
9	160.5	30.9	158	3	US-09-041-889-40
10	160.5	30.9	226	3	US-09-041-889-32
11	158.5	30.5	472	2	US-08-216-894-10
12	158.5	30.5	472	4	US-09-115-746-10
13	158.5	30.5	564	2	US-08-216-894-2
14	158.5	30.5	564	4	US-09-115-746-2
15	158	30.4	218	3	US-09-041-889-4
16	158	30.4	218	3	US-08-837-058-4
17	152.5	29.4	222	3	US-09-041-889-3
18	152.5	29.4	222	3	US-08-837-058-3
19	150	28.9	1507	3	US-08-929-329-5
20	144	27.7	434	1	US-08-097-830E-3
21	144	27.7	434	2	US-08-456-112B-3
22	143	27.6	214	3	US-09-041-889-27
23	141	27.2	100	2	US-08-460-890A-62
24	141	27.2	100	3	US-08-167-641C-62
25	141	27.2	100	4	US-08-460-971A-62
26	141	27.2	100	4	US-08-462-040-62
27	133.5	25.7	212	3	US-09-041-889-1

28	133.5	25.7	212	3	US-08-837-058-1
29	133	25.6	433	1	US-08-346-849-2
30	133	25.6	433	2	US-08-293-284A-2
31	129.5	25.0	113	3	US-09-041-889-14
32	129.5	25.0	113	3	US-08-837-058-14
33	129.5	25.0	147	3	US-09-041-889-13
34	129.5	25.0	147	3	US-08-837-058-13
35	129.5	25.0	220	3	US-09-041-889-2
36	129.5	25.0	220	3	US-08-837-058-2
37	128	24.7	262	1	US-08-403-379A-1
38	128	24.7	262	2	US-08-929-414-1
39	128	24.7	263	2	US-08-557-309B-51
40	128	24.7	263	3	US-08-834-306-51
41	128	24.7	263	4	US-08-993-674A-51
42	128	24.7	263	4	US-09-256-976-51
43	126.5	24.4	56	4	US-08-993-008A-6
44	124.5	24.0	1561	3	US-08-894-017-23
45	123.5	23.8	516	2	US-08-762-106-8

ALIGNMENTS

RESULT 1  
US-08-460-890A-64  
; Sequence 64, Application US/08460890A  
; Patent No. 5994109  
; GENERAL INFORMATION:  
; APPLICANT: Woo Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,890A  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/167,641  
; FILING DATE: December 14, 1993  
; APPLICATION NUMBER: 07/855,389  
; FILING DATE: March 20, 1992  
; APPLICATION NUMBER: PCT/US93/02725  
; FILING DATE: March 19, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 212/066  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single





```

RESULT 6
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-205-426-201

Query Match 34.3%; Score 178; DB 4; Length 223;
Best Local Similarity 54.5%; Pred. No. 2.2e-08;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;

QY 1 AKYAKK--AEKAYAKKAAKE---KKAYAKKEKAYKAAAEAKKAEAKKYAKAEAK 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 ARKAACKAPAKCAAKAAKAPAKKAPAKKA-ATRAAPAKKATAAKKAPAKKATAAKKAP 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 56 AKKEAYKAEKYYKAAAEKYEVA--AAEKKAEEAKKAYKABAAKAAK 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 AKKAPAKKAATKAAAPKAPKAPAKKAAATKAAAPKAPAK--KAPAKKAPAK 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399

```





```

Matches 51; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

QY 1 AKKYAKAEKAYAKAAKEKKAYAKAEKAYKAAEAKKKAAEAQKYA----KEAACA 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 AAEATKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEA 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 57 KK--EAYKAEAKKYAKAAAEKKEYA----AAEAKKAAEAAYK-AAAACAAAE 105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 TKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEA 430
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-08-216-894-2
; Sequence 2, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-2

Query Match 30.5%; Score 158.5; DB 2; Length 564;
Best Local Similarity 44.0%; Pred. No. 2.4e-06;
Matches 51; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

QY 1 AKKYAKAEKAYAKAAKEKKAYAKAEKAYKAAEAKKKAAEAQKYA----KEAACA 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 AAEATKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEA 490
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 57 KK--EAYKAEAKKYAKAAAEKKEYA----AAEAKKAAEAAYK-AAAACAAAE 105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 TKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEA 546
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-115-746-2
; Sequence 2, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI

```



THIS PAGE BLANK (USPTO)